Lauri Kytömaa

Road lighting user experiments - a review and a new assessment method

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 12.11.2014

Thesis supervisor:

Prof. Liisa Halonen

Thesis advisor:

D.Sc. (Tech.) Eino Tetri

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Lauri Kytömaa

Title: Road lighting user experiments - a review and a new assessment method

Date: 12.11.2014 Language: English Number of pages: 8+122

Department of Electrical Engineering and Automation

Professorship: Illumination engineering Code: S-118

Supervisor: Prof. Liisa Halonen

Advisor: D.Sc. (Tech.) Eino Tetri

This master's thesis reviewed the current, and notable past experiments that have involved road users, their preferences, both real and perceived safety, and the methods used. The literature review sought to identify the most influential attributes identified to date, excluding the conventional and technically measurable photometric light qualities. After the responses which could be directly predicted from the measures of luminance and uniformity of lighting, the subjectively assessed naturalness has been identified in several studies as the most important factor for predicting the overall appreciation and perceived safety afforded by the lighting. Naturalness as the highly influential quality was found to be also supported by other research, both by those which had sought to explain the perceptions of the environment, and those which had studied the psychology of personal space.

The review revealed the fact that there has been no attempt to build a comprehensive model for predicting the users' affective responses to outdoor lighting, or to the changes and improvements that could be made. Tests have assessed numerous properties of the lit environment, but there is no conclusive data of the connections between those proporties, perceived safety and general appraisals.

The reviewed studies unanimously agreed that more studies are needed. Also, emerging LED-luminaires necessitate a larger number of test installations for measuring user satisfaction. Therefore, a device was designed to help test subjects give their opinions faster and easier in outdoor test areas, as opposed the long questionnaires used in past studies. The output of the device was then validated, shown to produce consistent and valuable data of the test subjects' opinions. Finally, the results of a usability test were described, showing that the system could be used as is, or it could be further developed for use in a Living Lab envinronment.

Keywords: Road lighting, user experiments

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Lauri Kytömaa

Työn nimi: Tievalaistuksen käyttäjäkokeet - katsaus ja uusi syöttölaite arvioinneille

Päivämäärä: 12.11.2014 Kieli: Englanti Sivumäärä: 8+122

Sähkötekniikan ja automaation laitos

Professuuri: Valaistustekniikka Koodi: S-118

Valvoja: Prof. Liisa Halonen

Ohjaaja: TkT Eino Tetri

Tievalaistusta rakennetaan käyttäjien tarpeita varten, joten tarkka tieto valaistuksen vaikutuksista käyttäjiin ja heidän arvioihinsa siitä on tarpeen. Valaistus ja sen laatu vaikuttavat moneen piirteeseen ihmisten toiminnassa, arvostuksissa ja kokemuksissa, joten käyttäjien reaktioita täytyy tutkia edelleen.

Tämän työn kirjallisuuskatsausosassa selvitettiin ja arvioitiin käyttäjien kanssa toteutettuja ulkovalaistustutkimuksia, niiden menetelmiä ja merkittävimpiä löydöksiä. Useimmissa tutkimuksissa koehenkilöiden lukumäärä ei ole ollut riittävä luotettaviin tilastollisiin päätelmiin, ja osallistujat ovat arvioineet kymmeniä erilaisia subjektiivisia mittareita. Vaikka mitattavat käsitteet valaistuksen kirkkaus ja tasaisuus ovat merkittävämmät ennustajat valaistuksen ja sen luoman turvallisuudentunteen arvioinnissa, tarkasteltujen tutkimusten ja niiden taustalla olevien käsitteiden pohjalta löydettiin perusteltu ehdotus käyttäjien turvallisuudentunnetta eniten ennustavaksi kysymykseksi: tienkäyttäjät kokevat turvallisimman tuntuiseksi valaistuksen, jota he kuvaisivat luonnolliseksi.

Kirjallisuuskatsauksen perusteella käyttäjien subjektiivisten arvioiden ennustamiseen kattavasti kykenevää mallia ei ole vielä edes ehdotettu, vaikka lukuisia erilaisia subjektiivisesti arvioitavia ympäristön ominaisuuksia ja niiden suhdetta valaistuksen mitattaviin arvoihin on tutkittu.

Koska aiemmin käytetyt, testireittien lopuksi täytetyt kyselylomakkeet on koettu työläiksi, lopuksi kuvattiin uuden paikkasidonnaisten mielipiteiden kyselyyn tarkoitetun syöttölaitteen suunnittelun lähtökohdat ja tehdyt valinnat, sekä laitteen käyttäjätestin tulokset. Laitteen käyttö arvioiden kirjaamiseen kokemuksen hetkellä todettiin perustelluksi myös muistiharhan pienentämiseksi. Lopuksi osoitettiin myös, että järjestelmän tuottamista tallenteista voidaan tunnistaa kohtia, joissa valaistus ei ole tyydyttävää, sekä näihin kohtiin liitetyt kommentit.

Avainsanat: tievalaistus, käyttäjäkoe

Preface

My sincerest gratitude goes to the staff of Lighting laboratory, for all the constructive suggestions and supportive atmosphere. Special thanks shall be addressed to Prof. Liisa Halonen and my supervisor Eino Tetri for this opportunity and the interesting topic. I must also acknowledge Salu Ylirisku for the original idea of the input device, as without his idea this thesis would not have been possible.

This thesis is a part of the larger multi-disciplinary research project Light Energy, which belongs to the Aalto University Energy Efficiency Research Programme (AEF).

Otaniemi, 12.11.2014

Lauri Kytömaa

Contents

Symbols and abbreviations

Symbols

λ	wavelength
$V(\lambda)$	photopic spectral luminous efficiency function
$V'(\lambda)$	scotopic spectral luminous efficiency function
$V_{mes}(\lambda)$	mesopic spectral luminous efficiency function

Abbreviations

API	Application Programming Interface
ART	Attention Restoration Theory
EC	Europeant Comission
CCT	Correlated Colour Cemperature
CEN	Comité Europen de Normalisation, European Committee for Standardization
CIE	Commission Internationale de l'Eclairage, International Commission on Illumination
ESM	Experience Sampling Method
HPM	high pressure mercury
HPS	high pressure sodium
ISO	International Organization of Standardization
k-NN	k-nearest neighbors algorithm
LED	light emitting diode
LL	Living Lab
LPS	low pressure sodium
MH	metal halide
PRS	Perceived Restorativeness Scale
RVP	Relative Visual Performance
SPD	Spectral Power Distribution
TNO	Toegepast Natuurwetenschappelijk Onderzoek, Netherlands Organisation for Applied Scientific Research

1 Introduction

Lighting research has utilized user experiments since the dawn of the profession, and user experiments helped to establish the major criteria for road lighting standards in the 1950's and 1960's, when de Boer claimed in the results of several studies, that the required road surface luminance level should be at or near 2 cd/m^2.[1] The same value was suggested prior to that by Dunbar, but then it was based on a task with little relevance to traffic situations.[2]

In the decades that followed de Boer's reports, the actual lighting recommendations had adjusted those first results to meet technical viability; by 1967, CIE recommended levels from 0.5 to 2.0 cd/m^2 for different roads, along with minimum uniformity requirements [1, p. 142]. Later Caminada and van Bommel set the minimum acceptable luminance level for pedestrians - among other criteria - based on test subjects' recorded performances and assessments. However, historically, users have been mostly involved in probes into the limits of visual performance in various situations; e.g. how does the detection distance depend on the light levels, or what conditions must the lighting installation fulfill to prevent excessive glare.

Traditionally, the need for outdoor lighting has been reasoned with increased safety on the road. Anyone participating in the traffic environment can avoid possible conflicts more easily, when they can see the infrastructure and all other road users from farther and with a greater certainty. In all foreseeable conditions, an accident can not happen without a preceding traffic conflict. In 1977 in traffic safety researchers' workshop participants agreed on a definition of a "traffic conflict", and the definition has often been repeated in later research:

> A traffic conflict is an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged.

A trivial deduction follows from that definition: when the involved road users identify the conflict, and change their movement promptly and suitably, then that risk of collision is removed. The literature review in chapter 3 will demonstrate, that whether they can identify the conflict, depends not only on their sensory performance, but also on their momentary attentional capacity. Likewise, whether they can make the correct changes in their trajectory, depends on the sensory feedback, reflexes and their momentary processing capacity. Chapter 3 will also show that those parameters are influenced by the street lighting in the area.

The physiology and the biology of the human eye has been researched extensively, and we know quite well how the light incident on the pupil transforms to neurological signals. The human eye can adapt to luminance differences from sunny outdoor scenes with $3 * 10^5$ cd/m^2 to unlit night time environments, even down to $7.5 * 10^{-7}$ cd/m^2. The eye has two different kinds of image forming photoreceptor cells to handle the range this large, rod cells and cone cells. In adaptation luminance of under 0.001 cd/m^2, only the rod cells are operational and this is called scotopic vision. At adaptation luminance of over 5 cd/m^2, the photopic vision range, the cone cells' activity dominates the vision. Between those limits, from 0.005 cd/m^2 to

5 cd/m^2 both cells are active, and this range has been adopted as the mesopic range in the recommended system for mesopic photometry.[3, p. 67] The two cell types have different spectral sensitivity functions. The result from this is that the visual performance, the signals reaching the brain, and the experienced sensations are gradually changing when the lighting conditions move through the mesopic range.

Advances in lamp and luminaire technology have made it possible to also compare other attributes of the lighting, when in the past decades the only possible solutions were those that minimized the installation costs or the life cycle costs at a given minimum acceptable overall illuminance level. Already in 1984 the Finnish national guidelines for street lighting had provisions to lower the luminance levels when traffic was low, or to turn the lighting completely off in some conditions to save on operating costs. At that time, the only luminance level cited as scientifically validated was the previously mentioned 2.0 cd/m^2 [4, p. 10]; de Boer's experiments in the 1950's were given as the source. Such power saving methods and goals were not mentioned in a relevant master's thesis from 1960, but the same thesis claims a national lighting standard didn't yet exist, and that the illuminance levels that would be recommended were based on de Boer's studies, and unidentified Dutch studies [5].

It should be noted in this context that people have sought energy savings from lighting installations for centuries, even if the implemented actions weren't reasoned with any effects pertaining to the goals set for the lighting. Already in 1893 when oil lanterns had not yet been replaced by incandescent lamps, there were, even in Finland, cities with protocols to turn off the lights in the middle of the night [6, p. 70], apparently to save in fuel costs, and in 1908 the city of Jyväskylä decreed to keep street lights off on clear weather full moon nights [6, p. 83]. The relation between the perceived work required to generate firewood, and the light released by burning, can mean people may have a strong, learned, idea of the power lost when lighting is increased. At least compared to other uses of energy, we perceive the power more directly.

There has been an increasing interest in the psychological and physiological effects of street lighting: the light does not only allow us to see the road, but the studies reviewed in chapters 3.3 to 3.6 have shown that the light also influences the way we see the environment, how we react to it, and how it affects our emotional state. From usability research we know that the most practical, if not the best way to assess the effects on peoples affective responses is to ask the people, despite the methodological challenges. The alternative method, analysis of vast samples with advanced statistical techniques, is also plausible, but even with a very specific research question the data collection step could be a deterring factor.

The use of user satisfaction in assessing real world lighting installations has emerged to coexist with the conventional quantitative technical criteria used to date. Even the technical criteria are evolving, as the formulae used for predicting glare are revised (for example [7, 8, 9, 10]), and as we learn more about the human vision in the mesopic range, both of which require studies on the human end of the whole lighting system. It is therefore highly likely, that the need for both quantitative and qualitative data about user preferences is increasing while we reassess and refine the commonly used models.

Even if the laboratory experiments used to study for example detection thresholds or reaction times can be viewed as testing with users, this work will focus on the user experiments conducted outside, with real world lighting installations, be they experimental or industry standard.

Outdoor lighting does not exist in isolation. Societies have started reviewing the possibilities to enhance the user experiences of outdoor public spaces with carefully designed lighting, beyond the minimum measurable technical criteria used to date. As other fields of study, e.g. environmental psychology, architectural studies and human cognition research, reveal recommendable and positive features of favourable environments, the lighting profession must evaluate and embrace their findings to find the optimum and cost-effective solutions.

It can be argued that it is impossible to build an all-encompassing formula to predict the performance of any given task under any given lighting conditions. As Boyce puts it, that is because "no task is purely visual" [11]. However, contrary to what Boyce wrote elsewhere in said paper, later research [12, p. 54],[13, p. 95] has pointed out that lighting can affect the attention capacity, and thereby also the cognitive component of any task. Likewise, such formulae predicting the perceived safety and acceptability solely from the relevant luminance values will be an attempt in futility. This can only mean that a plethora of user experiments are required to assess the intermediate findings, and required for the road ahead – when we can't have a specific model that models everything, we can still have solid knowledge of what attributes and features have a positive impact, and we can have knowledge of the relationships between those attributes.

1.1 Challenges in user studies

There are a lot of pitfalls to avoid when users, their experiences and their preferences are to be measured and quantified for the purposes of finding the optimal solutions - and this applies in any area of endeavour. User questionnaires are often used as a simple tool for gathering users' appraisals, yet the questionnaire itself should be proved to provide reliable and valid data. For example, in software usability research there are some heavily researched questionnaires, for which the results are known to correlate with the actual attribute being measured. In lighting studies (beyond few comprehensive studies) to date only the lesser content validity of questionnaires has been addressed, if at all. Only a select few studies reviewed later in this thesis addressed the construct validity of their measures.

With regards to lighting and the users, previous studies have identified, or at least speculated, a long list of factors that can affect the appraisals in one or both directions. It is also known that some of the parties involved might not notice the effect of those factors. Beyond the questions of reliability and validity of the questionnaire itself, the list of factors to consider includes environmental factors like

- The time of day [12, p. 12], by itself and in relation to the chronotype of the test subject [12, p. 27]

- The ambient temperature, both in isolation, and compared to the correlated colour temperature (CCT) of the luminaires used [13, p. 99]
- The weather and climate effects; for example snow cover [14] or presence of foliage on trees [15, p. 34]
- The common quantifiable metrics of the lighting setup, e.g. distribution and different illuminance level measures
- The presence of other people [16]
- The common knowledge of safety issues in the area [17, p. 199]
- The classification of the area as urban/suburban/rural [18]
- Whether the area is considered public or private [13, p. 55]
- Lack of knowledge about the real salient features in the scene[13, p. 99]
- And other variations in the scene beyond the task area.

Also things influencing the test subjects' cognition need to be assessed:

- The knowledge of the test's purpose can affect their attention [13, p. 98]
- Any descriptions of the values, or the study questions wording can have a big impact, either by being ambiguous or otherwise assessed from different cues by different subjects [19, p. 288], or by affecting the test subjects' perception of the attribute they were asked to assess (semantic priming)
- The descriptions of the scores and questions can mislead their thoughts into fear they would not have felt otherwise
- Whether the subjects are local to the area, or in an area unknown to them [13, p. 80]
- Their mental fatigue, and stress levels before the study [13]
- Their motivations and goals, both in the study and in the area in general, e.g. relaxation or cycling [20, p. 47]
- Cultural factors, i.e. the learned expectations and patterns of conduct [21, p. 325–327]
- Age and other personal variations (e.g. [22, 17])
- Hindrances and difficulties the test subjects experience in stating their preferences [23, p. 27]

Finally, also study setup related factors can influence the results:

– Whether static images were used, or video, or real environments [24]

– Accommodation period length [25]

– The extremities of the scale used in the study (range bias) [26, p. 410]

– Whether the question's answer is more likely yes than no - in some cases, when people are unsure, they are more likely to answer yes than no (response bias) [27, p. 333]

– The number of values in the scale [28]; e.g. contraction bias, where an odd number of choices in a Likert scale is known to emphasize the midpoint in the answers [26, p. 410], but it is often more important to have the neutral choice available. Also grouping bias, where test subjects have to group a large number of different stimuli levels into a single category so that the grouping reflects the ease of grouping, and not the actual relative levels of the stimuli.

– The initial or reference value in a comparison or adjustment test [29]

– With repeated assessments, after the first assessment the next responses are affected by the range of previous stimuli [26, p. 410]

– Measures to mitigate position bias, interval bias; these include counterbalanced ordering and including null condition tests[30]

When conducting user experiments, one must always remember that the study setup must be carefully designed to overcome common mistakes with subjective answers. In previous studies, it has been shown for example that the initial lighting level and the range of lighting levels available in the experiment affect the choices made by the test subjects [31, p. 404]. That is to say, that the results are different even if the only difference is that the minimum or maximum dimming level is artificially restricted (range bias). The research conducted by Logadóttir et al. [31] to the effects of range bias in lighting was about indoor lighting at luminance levels higher than the ones utilized in road lighting, but the same effect can be one of the reasons why several countries implemented different reference road lighting levels [32] in the 1960's and early 1970's.

Some of the bias effects listed above are not an issue in outdoor user experiments, or they might be unavoidable, depending on what the research question is. For example in the laboratory test of assessing glare from nonuniform reflected light patterns, most of the subjects had a strong personal bias to always indicate the left or right side test signal as more discomforting, when the signals were identical [33, p. 205]. Still, the effects need to be addressed when designing the study.

All in all, because of the great number of variables, and possible complex interactions between some of those variables, a significant number of different studies still need to be conducted to validate the findings of past studies, and to form a comprehensive picture of the most significant interactions between outdoor lighting and the preferences, actual and perceived safety, and perceived restorative capabilities. It is therefore beneficial to study how the preference data collection could be faster, less intrusive and easier.

1.2 Outdoor lighting future directions

There have been a number of studies into the feasibility and benefits of smart control of road lighting, i.e. systems that change the lighting parameters dynamically based on road usage and other sensor data. The conventional motivation for a more complex control system is the possibility to save money with reduced energy consumption. One should not, however, bypass the escalating evidence of the harmful effects of exposure to light in the part of the circadian rhythm where darkness is expected, or the evidence of the malevolent changes in plant and animal behaviour near lit areas (for an overview of the effects and the causes, see the review by Navara [34]). The negative effects are not limited to animals; extraneous light also affects the human circadian system and our mood.[12, p. 42–45]

The estimates of achievable power savings are in the 20 % to 50 % range, and with an aggressive dynamic dimming even 77 % has been suggested [35, p. 39]. Such savings could be achievable first and foremost by dimming road sections without any users, but also by reacting to weather, climate and the dimmest daylight conditions. Depending on the length of dusk and dawn, considerable savings could be realized by turning the lamps on and off gradually and only at times adapted to hyperlocal ambient levels.

In areas with long twilight, i.e. at high latitudes, old clock timer or central skylight measurement based control systems easily keep the street luminaires on for 30 minutes longer than necessary, at both ends of the dark time. Instead of using full power for the time interval based on the worst case scenario in the area controlled by a single control device, in one study in Helsinki the luminaires would reduce their power consumption by 62 % during twilight, without compromising the horizontal illuminance level [36]. The literature review found no studies about the users' preferences in those partially dark times.

In foul weather, the required lighting levels increase, but for example snow covered ground is much more reflective and without dimming the attained luminance levels are roughly fourfold higher than on a dry road in summer. The weather does, however, also affect the other characteristics of the perceived lighting, especially the uniformity is decreased in wet conditions. The high reflection factor for specular reflection, i.e. concentrated in one direction, might even result in discomfort, where the wet road surface has next to a lower luminance area a high luminance reflection. [37, p. 95]

Some experiments have been made about the dynamically controlled road sections, i.e. the acceptable dimming parameters in relation to the traffic volumes and weather conditions have been explored [38]. However, no reported studies of the road users' perceived acceptability of such real world dynamic dimming systems could be found; such appraisal tests have employed short test installations only. The system designs have considered human perception, though, by limiting the rate of change of the luminous flux in such a way that the road users should not be able to observe any change [39].

In the recent years, LED based light sources have become a viable alternative even in many outdoor settings. One major benefit of LED's is the fact that their dimming level can be varied instantly, whereas conventional light sources capable of producing the luminous fluxes required for outdoor lighting usually take even several minutes to reach the intended level. Research has also shown that the LED luminaires can achieve a better colour reproduction than the high pressure sodium (HPS) or mercury vapour lamps. Despite the open questions in colour reproduction under LED lighting (for an overview, see [40]), it's well agreed that the currently defined colour rendering indices do not reliably predict the users' acceptability of white LED lights [41, p. 53], [42]. Research is ongoing, and practical uses will require user experiments with various luminaires in the years to come.

1.3 Research methodology - past and present

The practical lighting research is to some extent dependent on the technical measurement instruments available. For example, the imaging luminance meter has vastly sped up the ordinary uniformity and mean level studies (compared to spot metering multiple times); with more measurement points the final values can be better indicators of the underlying visual performance level, which the mandated minimum luminance level tries to ensure. In laboratory based tests the automated test methods have allowed a greater number of samples for each test subject - and better test designs meant to eliminate unwanted biases - without significantly increasing the time required.

In the user experiments outside the laboratory, the technological advances have mostly contributed to measurements of isolated vision related task, e.g. recording detection thresholds and reaction times. In surveys into the test subjects' preferences or about their social behaviour, no such notable steps toward faster and more efficient data collection have been realized. That is not to disparage the tool improvements unrelated to lighting research itself, mostly realized through common office software.

The idea to monitor test subject's actions and reactions in the real life with portable recording devices is not new, but still some years back the devices would have lacked in at least one of the critical attributes: less intrusive methods, sufficient battery life or miniscule costs. Already in 1982 there was an experiment with a handheld battery powered reaction time testing unit, built into a cassette player; the device recorded the responses onto the cassette, which could then be analyzed later [43].

Several studies have utilized computer simulations of outdoor environments ([13, 44, 45]) and others have assessed the validity of the test subjects' responses to such simulations[46, 47, 48]. A comprehensive validity study by Bishop and Rohrman [46] came to the conclusion, that both in the real world and in the simulations the differences between daytime and nighttime are similar, but nevertheless the responses are not the same. It might be inconsequential that their study did not report any comparison between the illuminance levels of the presented simulation and the real environment, as they refer to the fact that the environments are "experienced within a subjective context". That is to say, that the subjects use their prior knowledge of the environments familiar to them when assessing the newly presented environment, and can somewhat compensate for the shortcomings of the simulation. Later, Newsham et al.[49] paid considerable attention to the spot illuminances of their indoors simulation and their results showed better correlations between their simulation and the real indoor scene it depicted.

In turn, it is easy to claim that for new technology light sources (like LEDs) and for new dynamic lighting control situations, even the simulations could produce reasonably accurate responses when the same area is compared to a lighting setup with conventional light sources, but the results would need validation with actual installations. At the moment, because we don't have validated models for the effects of the special characteristics of said dynamic lighting control schemes, or models for the effects of LED luminaires on the perception, actual test installations are much more valuable.

1.3.1 User appraisal of street lighting

The first historic ideas in street lighting, that measuring only visual performance would be sufficient for practical installations, has long since been abandoned: the lighting also has to be perceived as positive. Then one has to ask, how to measure the perceptions of lighting?

When examining the effects of light on humans and designing the experiments, one should remember that there is no placebo for light [12, p. 28], [50]. In other words, this means that to separate the direct effects of light from the perception of light, when that is necessary, the study method needs other tools beyond sample comparison.

To date, various researchers have used a plethora of questions and answer scales; these are explored in more detail in chapter 3. Sometimes the questions have been taken from prior studies, sometimes researchers have applied terms and concepts from other fields of study, namely environmental psychology.

The first group of applicable questions can be derived from lighting research's own findings: for example, we don't have affirmative, universal equations for predicting glare in all conditions – discomfort glare being a subjective experience – so we can ask the subjects to rate the glare. Likewise, we can ask, say, how much they prefer light with CCT of 4000 K, and compare to appraisals of light of another CCT, other variables being equal. These are trivial to come up with from the technical and physical measures of light, and only studies can reveal which of them are actually useful for the visual comfort.

The next group of questions can be derived from the reasons to install lighting. If we light up the streets, there has to be a benefit to spending money on the installation and running costs. The main objectives of street lighting are the traffic safety, improved traffic flow, public security, accessibility and perception of space [51, 52]. Some of these can be turned to questions trivially, but for example public security can be broken to many different aspects, from a perception of lack of crimes to identification of vandals' identities, and to the motivations behind the people's actions in that space. For meaningful questions pertaining to these points, the concepts need analysis of their theoretical background. Such analysis has lead to, for example, the now common questions of prospect, entrapment and concealment;

these are explained in chapter 2.5.2.

1.3.2 Living Lab

A setup gaining popularity in multi-party innovation research is the Living Lab concept (LL), where several stakeholders collaborate to empower the users to contribute their ideas and feedback in real-world settings and contexts. Currently, the term Living Lab is used in various meanings, all emphasizing different aspects and benefits of testing and innovating with real world users in their everyday environment. A Living Lab is not solely for scientific research, nor is it solely for solution design.[53, p. 11–12]

A LL tries to overcome some of the limitations in prior design methods involving users, which are related to the fact that the methods come from an organizational background. For example, the User Centered Design approach is based on the traditional workplace context and processes, and has problems transferring continuous interaction between different contexts.[53, p. 3]

The LL is not a single method, but a methodological approach that combines several methods, with features of lab reseach and real world action research. The actual methods in use can be positioned on two dimensions. The first axis goes from laboratory (conductor has full control of the situation) to "living" (no control), and the other axis represents the mediation: by information and communication technology, or by the researcher. In a LL, different methods are used both concurrently and consequtively. Users, their actions and experiences, need to be analyzed in different context, and to know the context, one has to know something about the context, from the location to the users' social interactions at that time.[53, p. 63–65]

One explored direction in this thesis is the possibility to use an existing Living Lab to gather the users' subjective assessments related to outdoor lighting. One of the required tools to a successfull LL is the capability to handle larger and more dispersed datasets than normally used in design methods: this can be compared to the crowdsourcing efforts online, where all problems can be considered easy given a large enough number of users participating [54]. A successful LL however adds an innovation force by uncovering the hidden needs of users, and unexpected user behaviour patterns from their experiences.[53, p. 15] Although software and online communication of data are a crucial part of the LL, the users in LL are not using distinctively online services, but they are interacting with their real-life environment.[53, p. 19]

User involvement in a LL is not static, but goes through distinct stages: community development, definition of interest areas collaboratively, eliciting users' needs, encouragement of user participation and evaluation of results together with forming of initiatives for future collaboration. A successful living lab requires different stakeholders to adhere to the best practices and that they evaluate the effectiveness of the decisions so far.[55] The descriptions for such processes and methods are yet to reach maturity.[53, p. 6]. Although the service development and innovation in a LL are said to be fast and realizable with relatively little effort, the required networked system of different service providers has several layers, services and modules, which have to be customized to each Living Lab. [53, p. 31] This is a natural consequence of the involvement of various stakeholders and of their requirements. That complexity is also a weakness, something that can hinder the project's completion, when one is designing their LL. It is then evident that a a Living Lab is not a lightweight undertaking. Especially if the planned LL is small, the overhead can be considerable when compared to older research methods, but especially with larger studies the LL approach can improve efficiency.

There has been at least some studies that explored how an adaptive lighting affected the acceptability of the lighting on a short road section[56]. However, the results were not sufficiently detailed to allow extracting any kind of theory hypothesis, or a working model. It is reasonable to argument that because of the large network of complex interactions, it is unfeasible to create a universal vision-cognition model by conventional research methods alone. At the very minimum it would require an extensive database of users' appraisals, only realizable through crowdsourcing in a LL environment. Even if we have less exaggerated goals in user experience research, a fully developed LL environment would provide useful co-operation and data of the outdoor experiences of lighting.

1.3.3 Crowdsourcing

"Given enough eyeballs, all problems are shallow" is an oft cited extension of the Linus' law, which was first coined by Eric S Raymond in his article on the then emerging massively distributed software development method. The crowdsourcing method can only be successful, when the project initiator successfully provides a working example, which convinces others to see the potential in it evolving to something really neat in the foreseeable future. In this "bazaar" coding style each contributor only has to fix a small detail, one at a time, in the context that matters the most to them personally. (The conceptual opposite is the "cathedral" method, where each change and allocation of resources is planned ahead by the few in charge of the project.) [54, p. 19 & 44]

Raymond had used the word "bugs", referring to mistakes in the code, instead of "problems", but it some years later it was realized that this appears to also apply to other collaborative projects. The term *crowdsourcing* was first defined in 2006 by Howe in an article [57] in the *Wired* magazine, describing a strategy of outsourcing to the crowd, a strategy which was already used at that time. Originally it was only used in the online Web context, and the process is most easily manageable online, but Howe's wording didn't really impose any such limits. In crowdsourcing, the initiator posts an open call to a large network of potential labourers, who submit vast amounts of solutions, which the initiator then aggregates and utilizes for their own gain; the winning solutions might be rewarded.

In outdoor lighting research, a crowdsourcing viewpoint would take subjective assessments from thousands of users and up to millions of data points, and use statistical methods to extract the good and bad installation locations. The conventional method for studies of lighting installations in contrast would have a set test route where willing participants agree to evaluate their impression of light at the given locations only.

There is some research pointing to the "wisdom of the crowd" at least in specific contexts, i.e. the aggregate of the answers of a large group of people can offer solutions almost as good as the best solution any of the individuals in that group proposed.[58]

In street lighting research, this is something that remains to be tested. Such a study would have, for example, participants wield an input device throughout their day-to-day activities, and at their leisure give feedback of the lighting at their current location. With the large network of participants, the great number of assessments tied to an identifiable location could then reveal areas with exceptionally bad (or good) lighting. Depending on the design, the reasons for the assessments could either be collated from open ended comments, or the locations so identified would have to be assessed with expert reviews and localized studies. The device presented in chapter 4 is one first step towards such a method.

Crowdsourcing has its downsides, too. Brabham [59, p. 83] offers an overview of the plausible negative effects of crowdsourcing on the crowd, from unpaid working hours for the unsuccessful solutions to loss of livelihood for professional experts of the field in question. However, such major problems should not be an issue in the proposed lighting acceptability research method. The only issue to address is the availability of the required technology, and the usability of the device, so that the contributing user base is a representative sample of the whole population.

1.4 Research question

In the previous section the missing links between known studies, present technical design standards and known future challenges were identified. The research for this thesis was conducted within the Light Energy project and was set to provide value for two of the project's main objectives: to develop tools to assess the quality of visual environments, lighting and traffic services, for the safety and conformance to users' needs; and to facilitate the collection of data on the user experience of outdoor lighting in a Living Lab.

The research question is then twofold:

1. Based on a literature review, evaluate and select the subjective parameters that have the greatest support in previous research as having the greatest significance on the road users' actual or perceived safety, and on the perceived quality of the street lighting

2. Develop and validate an input method better than end-of-session paper questionnaires to store those subjective assessments in a real world environment

In the context of this thesis, users' perceived safety always refers to a consciously evaluated assessment. That is to say, that in this context we are not interested in the possible unconscious physiological changes that might correlate with, cause, or happen because of an escalation in apparent danger, nor in the actual risk level of the area. It is the conscious fear that affects people's choices and behaviour [13, p. 55].

In the constructive development, the new method shall be evaluated as better if the produced data is consistent with the models possibly identified in prior research, and if the input method requires less effort on any parties present in an outdoor lighting appraisal test.

1.5 Outline of this thesis

This chapter has provided an introduction to the concepts, phenomena and methods relevant for the research question of this thesis.

The next chapter 2 introduces the relevant theories and methods for understanding the findings and methods of prior studies reviewed for this thesis. These include both the known effects of lighting on human cognition and perception, and also the challenges in conducting research with test subjects. In addition to that, the chapter provides more detailed reasons for assessing the perceived safety and affective responses, and assessing the effect of lighting on visual performance.

In the literature review of chapter 3 the prior studies somehow relevant to the research question are described first. For those studies that do not seem trivially related to the goal of this thesis, the coined reasons for inclusion are listed. In the last subchapters their methodologies, research questions, concepts and findings are discussed to answer the first research question.

The chapter 4 first describes the technical and user interface requirements for the research device developed in this thesis, with reasons founded in the findings of the prior chapters. Then chapter 5 details the validity tests conducted with the device.

Finally, the discussion chapter 6 reflects the coined requirements and the design choices against the previous knowledge.

2 Background

Outdoor lighting installations seek to provide the optimum minimal acceptable illumination levels with lowest possible costs, yet suitable for the needs of the users of the area. The luminance levels outdoors are generally in the mesopic range. Indoor installations typically utilize even several magnitudes higher luminance levels and the purpose of the light is to improve other functions and perceptions; for a building operator increasing the lighting levels (or, specifically, generally better lighting) is an investment to get higher income and smaller overall costs. In outdoor locations, there is often no income to gain, and the reduction of other costs is harder to assess. This is because of the scarcity of traffic accidents, and due to the fact that the factual accident costs are dispersed among several parties, for example several road operators, insurance companies and private individuals. Also, in outdoor locations it is far more difficult or laborious to isolate the effect of lighting from other factors affecting the users and their behaviour, than it is indoors.

The models presented so far for predicting visual performance are remarkably similar, but they all incorporate a threshold under which the performance deteriorates unlinearly [11]. As a result, the results from indoor lighting studies with users often can not be directly applied in the outdoor lighting context without validation. Because of that, indoor user studies were mostly not included in this thesis.

The end users of lighting are involved in two different kinds of experiments in road lighting research. After this research of reported studies in the recent decades, it should be safe to claim that the majority of scientific studies conducted with users to date, have investigated the test subjects' ability to notice visual sensory stimuli in different parts of their field of view, or the reaction times to that test stimulus, and also the detection thresholds of the shapes, colours or luminance levels presented. The other type of end user testing is designed to uncover the test subjects' reactions and emotions prompted by the different lighting setups, no matter whether they are reactions to the detected test objects, or real life interaction. The review part of this work has an emphasis on the latter group of tests and their methods, because the goal of the development work in chapter 4 was to come up with a working measurement tool for those subjective appraisals and reactions.

Where road lighting is installed, it should not provoke a false sense of road safety, which can mislead the road users into misjudgments, for example to choosing a higher speed. There's research into whether that happens, and for example the Finnish Road Administration has reported that installing street lighting on a rural main road didn't affect the speeds, probably because the night time speed was already at the speed limit [63, p. 48]. Other studies cited by that report, in other countries, have found an increase of 3.6 km/h. This is more obvious in urban areas, where drivers more likely choose speeds below the limit, even if they do drive faster at night than they do at daytime [64]. That is to say, that it should be physically possible for the road users to see, in foreseeable traffic situations, the relevant objects reasonably early so that they at least have the possibility to react in time. The visual ability tests mentioned in the previous paragraph serve, and have served, the important purpose of setting the relevant minimum levels for different traffic environments. This has not been in isolation, but has drawn also from the research on human cognition, and the processing ability limits involved.

2.1 Usability and lighting research - the connection

Although usability professionals can use expert reviews to improve any systems design, their results generally come second to any usability tests conducted with the intended or actual users of the product. Because usability testing with users is the more laborious method, the tests must be adapted to the situation to minimize the number of test subjects without compromising the validity of the results. Taking into account the great number of factors affecting the user experience and vision threshold, as outlined in the next paragraphs, it is evident that numerous user studies will still be needed in lighting research.

The results of the user studies will help formulate lighting design guidelines, and will provide comparative data on different luminaires' perceived value to the users. The traditional questionnaire forms will still be in use, but for detailed probes into specific research questions, developing a faster and more lightweight data collection method would be beneficial.

In chapter 4 this thesis describes a new device that could be used as the input tool in such a research method. With the device, the user's appraisals of the most important factors affecting their experience can be collected in-situ while walking in the test environment. This should yield more a accurate presentation of their impressions and will be less laborious for all parties involved.

Usability expert reviews require that the users' tasks are first defined, so that the expert can evaluate whether it is likely that the user can succeed. Later, chapter 2.5.1 will explain that the possible situations in the traffic system number in the hundreds even when conflated - and that does not include for example pedestrian environments. These two factors combined lead to the conclusion that expert reviews of the user's experience of street lighting are next to impossible at the moment for two reasons: all of the relevant tasks are not known and the number of situations that would have to be considered is unworkable.

2.2 The concept of usability

Although some new products still have basic usability problems, the usability research as a profession improving the man–machine interface is not new. Before World War 2 there were first known systematic efforts to minimize operator errors in fighter planes, and research to the phenomena behind the errors. In the first decades, current day usability was a part of the more general human factors research. Usability, under that name, was only popularized starting in the 1990's and has since gained momentum with the increasing use of various user interfaces in our daily lives. Research on the lighting's effects on humans and of the effectiveness of the lighting is, roughly, as old as conventional usability research, evident from the de Boer's book which listed the primary reasons to install lighting as accepted knowledge [1]. User centered design was called for in academic papers already in

1967 by Knowles, as Poulton reports in 1976 [26, p. 409] that others prior to him had already suggested product design choices should be based upon the assessments of numerous users, instead of the design most efficient by technical measurements.

There has been a shift from strict usability to designing for the more broader concept of user experience. Improving the user experience considers, in addition to the immediate user-system interaction, the emotions and perceptions of the user, their past and future expectations and motivations, i.e. how the system influences those.

The definition of usability in the international standard ISO 9241-11 is widely accepted as the basis of usability research, even if there are other, but very similar [65, p. 26], definitions in active use. Before introducing the definition, it's worthwhile to note that it encompasses more than the immediate human-system interface.

The ISO 9241-11 defines usability as: [66]

> *Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.*

Originally the whole ISO 9241 standard was titled *Ergonomic requirements for office work with visual display terminals* and other parts of the standard still cover ergonomic aspects of mainly office work, from workstation layouts to worker's task requirements. The knowledge on which the standard is based on, is however in many places more general, and the part 11 concerning usability, as it defines usability above, has in it nothing that would limit the methods and definitions to only one type of work environment.

The usability can be assessed for any system involving users and interaction; such as lit streets. One should also note that the definition given in the standard does not define any measurable criteria or units, but the concepts applied must be adapted and remodelled to the context under scrutiny, before the usability, or changes in usability, can be measured in that context. Therefore, the definition can be dissected and reapplied to road lighting:

The product does not need to be a single object, but in this context it is the whole road traffic system including the road, the road fixtures and the lighting

Users are, literally, the users. It is noteworthy, that in urban conditions there are multiple, diverse user groups that have their distinct viewpoints, and different needs and expectations, just like users of other products.

The context of use is the sum of the different factors affecting the road user's route, their interaction with the system and with the other users, and the environmental conditions (such as weather)

The goal is mostly to reach their destination, but especially the pedestrian road users' goals can include mental restoration, relaxation, workout and other leisure related benefits.

The effectiveness refers to users actually accomplishing their goals. Any system can not be usable, if the goal is not reached. The measure for effectiveness can therefore be anything from reaching their destination to lowered stress levels, from relaxation to new experiences.

The efficiency refers to minimizing costs – monetary, physical and cognitive, and the time spent pursuing the goal.

The satisfaction is maximized when the interaction with the system can minimize the physical, mental and sensory load, and when the possible positive feelings are increased.

The overall satisfaction in the society can be increased or decreased because the lighting can have an effect on suitable land use (zoning) nearby. Most notably, this has been one factor in limiting the street luminaires' stray light above the horizontal plane.

We can not rely only on technical measurements when assessing the effectiveness and the acceptability of outdoor lighting. Also people's behaviour and their wishes have to be considered. Usability researchers and psychologists have studied the possible methods to measure those reliably.

Guidebooks to old lighting standards have listed the measureable values most appreciated by the drivers in order: evenness, luminance and lack of glare.[4] These are still used and easy in the sense that they can be measured without the users, but other criteria are still researched. These affect the satisfaction and efficiency components of usability; comfort is increased, and when observation errors and missed visual targets are minimized, the road users spend less effort in correcting their errors.

The definition of usability includes the context of use, which means, that if the users can't control it to their liking, the context must be considered when assessing the usability. In the usability tests of lighting conditions, this means that the tests should be repeated in varying weather conditions, unless further research provides reasonable evidence that the user's perception in all weather conditions can be predicted with the usability results obtained in some set weather. Other major factors to consider in defining the context are the subject's mental state (especially the stress level), their motives for accessing the area, the perceived safety of the area (by each subject) and the number of other persons present.

2.3 Usability testing methods

Although user interfaces and usability in general can be assessed with *expert review* methods, where a trained professional goes through the elements of the interaction system with the accumulated knowledge of prior usability reviews, and finds the most likely usability problems, the systems must be evaluated with the real users to find the real and the most serious problems for that particular user group and system. To get the best representative data of those problems, and to avoid common study errors when testing with users, different usability testing methods have been formulated. Their details are not set in stone, but the descriptions rather set the outline and each practitioner adapts the methods to the context of the inspected system and to their own goals. Only some of them can be reasoned to be plausibly relevant for outdoor lighting research.

User testing is the method most widely known by the public. The users are asked to perform predefined tasks, and their progress is monitored and later analyzed to identify the issues they encountered. Monitoring can include eye tracking and user's vocal descriptions of their intents and thoughts as they work on the system. In traditional user testing the researcher does not intervene in the user's actions, and has to be trained not to influence the test users' attitudes and perceptions. Usually the users are asked to fill a questionnaire about their perceptions about the tested system after the test is over. User testing can be performed on prototypes, and working products.

In outdoor lighting research, this is not unlike the tests already conducted with real installations, even if there are differences. To date, the researchers mostly have analyzed the recorded appraisals or actions after the test is over, whereas intermediate notes about the actions, reactions and affective responses during the test are missing. Only eye tracking studies and few distance to glare comparison measurements have implemented continuous recording.

The context of outdoor lighting poses a challenge for applying a test method more closely resembling user testing; the very presence of the researcher can change the test subjects affective responses, especially so in areas where there would not normally have been any other people nearby.

Pluralistic usability walkthrough is a method where system designers and users gather to see how the users try to work on the system or with the prototype, in an event managed by the usability expert. The designers work as a living documentation for the users if needed, but should not intervene in the user's progress through the use cases. The users describe their actions and solutions, and finally the results are discussed among all participants. The expert makes sure the designers do not try to reason away any problems encountered by the users. Benefits include that the designers get immediate feedback, new ideas from the users and even more motivation to work on the user interface.[67]

In outdoor lighting research, the method does not seem to offer significant value, because it is more focused on the actions and choices users make interacting with the system. Although there have been some tests with lights on demand by SMS systems, road users don't manipulate the lighting consciously.

Especially online services – not because of them being online, but because of the large number of users – can utilize the method called *A/B -testing*, where a random sample of users is presented with a slightly modified user interface. When their progress and task accomplishment rate is measured quantitatively, statistical analyses can be used to reveal the better design. The major benefit is that such small improvements can be made incrementally, even one every day when the number of users is high enough. With a high number of research subjects, this could be employed in lighting research, say, by alternating between two different dimming strategies on alternating days (or hours) for a longer period.

The *Experience Sampling Method*, as the name says, is aimed at evaluating the experiences instead of the momentary usability. It is best suited for studying long term changes in the user's appraisals, by ensuring that the users give feedback repeatedly, but at the very moment they are prompted to do so. This method is meant to overcome the memory effect, i.e. to prevent the respondent's memory from altering the answers. Most other usability methods can only study the users' performance and the acceptability of the system at one point in time. However, it's well known that with repeated exposure the users learn to use the system and that their internal state changes with time, which both change their preferences and performance with the system in question. Teague et al. [68] even demonstrated that the same system is rated, on average, differently for ease and enjoyment if the ratings are only asked after the task is complete, or if subjects give ratings during the trial; they suggested that when doing retrospective assessments, people de-emphasize prior negative events.

2.4 Experience Sampling Method

Not strictly a usability method, the Experience Sampling Method (ESM) has been used in social sciences and psychology to examine the quality of life – *what* people do and *how* they feel about it.[62, p. 3] Although it requires application of – often customized – mobile devices, the first studies where it was used were conducted already in the 1970's using the then emerging beepers. Technology advances have only facilitated cheaper and faster study setups.

In ESM, the participants are paged at random times to record a note of the phenomenon under scrutiny, usually their experience of it and the study is continued on time frames up to several months. The note can be for example a questionnaire with open- and closed-ended questions, freeform text, or a measure of their emotional state.

Experience as a concept includes many facets of the consciousness: feelings, sensations and thoughts. In relation to street lighting, this means that the experience of a street is different every time one travels there, but each experience also has common distinctions with other places and activities. ESM seeks to uncover these distinctions.

The notion of experience is built on psychologists models and approaches; in the phenomenologist approach the researcher does not need to know what happens at the level of physical interactions when, say, a street is relit, but they study how the persons there experience the change of lighting, how they describe the lightings effects to themselves or to others. The behaviourist approach on the other hand largely ignores how people perceived the lighting, but takes detailed notes of what the subjects did, how they altered their behaviour after the relighting.

In everyday life, people tend to analyse the social world with both approaches, with constantly varying emphasis on one or the other. The stream of events we perceive is then a sum of past and present allocation of attention, experiences. ESM tries to measure the changes in that stream over time.

By just observing what people do, the cognitive and affective dimensions of their experience are lost. Usually in ESM the miniquestionnaires submitted by the volunteers contain some questions to the physical world they're in at that moment, so that their answers can be tied, to some extent, to their actions.

The random timing of the question prompts avoids the effect of a fixed schedule on the subjects cognition, i.e. they don't alter their behaviour in anticipation of the upcoming answer task, nor do they have to set aside time to try to remember what happened some time before, as they would with retrospective diary studies. They do, however, need to employ some form of self-selection and selective nonresponse; the timing of the answer prompts can never be totally undisruptive.

As their answers get written down when they are experienced, the cognitive biases in the answers are reduced.[23]

A more comprehensive overview of ESM can be found in the book on the method, by Hektner et al.[62].

Another similar technique applied in user studies is that of lifelogging, where the participants carry a camera, and either consciously take photos of relevant features in their present environment, even hundreds of photos each day, or an automated miniature camera records their actions at fixed intervals. The number of photos can be excessive, especially with longer study durations and multiple subjects, such that the significant and recurring features need to be extracted from the photographs automatically.

In the context of street lighting studies, that method has some technical challenges: the lighting levels at night are low, such that the cameras in mobile phones, and most inexpensive and unintimidating stand-alone cameras operating in a point-and-shoot mode automatically utilize shutter speed so long, that the risk of blurred images is too high. Also the field of view isn't sufficiently wide, and therefore it would be hard or impossible for the researcher to later interpret the direction of the photos. Likewise, for example a photo of a street luminaire by itself does not tell us much about the environment around it. In conclusion, with currently available technology that method does not offer benefits over other data collection methods in the types of street lighting studies considered in this paper.

2.5 Vision, weather and safety

In outdoor lighting the environment and weather conditions have a considerable impact on the perceived lighting. For example, wet ground has a higher specular reflectance than dry ground, i.e. more of the light incident on the surface is reflected to a smaller space angle (whereas the diffuse reflection off a dry surface scatters the light to larger space angles), the wet surface exhibits larger spatial changes in reflected luminance, i.e. less uniform surfaces, and a large proportion of the surface can have a lower luminance in wet conditions.[37, p. 94–95]

As has been noted, safety is considered one of the main functions of street lighting, and the effect of weather conditions on measurable safety effects has been studied. Wanvik reported in 2009 in his analysis on the effects of street lighting on traffic accidents, which was based on Dutch accident statistics 1987–2006, that the installed lighting was more effective in good weather. In good weather, lit roads were safer than unlit roads; in rain, they were still safer than unlit roads, but the difference was smaller.[69] (N.B. rain increased accidents in both conditions.) The literature reviewed in this study did not mention any mechanisms, not contemplated nor studied, behind such interaction.

However, computer graphics and image processing research has made extensive use of modelling the physical world and the reflections and refractions present in common materials. Although studying methods to remove rain from real world videos, Garg and Nayar explain the physics that could explain the phenomenon [70], even if they don't mention street lighting: given that every raindrop refracts light from an approximately 165 degree space angle for each viewing direction, the luminous intensity refracted towards the viewer is essentially constant while the drop is in the field of view. They go on to explain, that this means that the background behind the raindrop at any moment has only little effect on the amount of light refracted through the raindrop. They continue that in normal weather conditions, the raindrops are sparse, i.e. the distances between the raindrops are large relative to their sizes, and because of that, the raindrops have virtually no effect on the luminous intensity of light refracted through the next raindrop; the total solid angle subtended by the environment is by far greater than that subtended by other raindrops.

The consequence of Garg's article, relevant to street lighting, is then that with this mechanism, rain drops diminish the contrast of the whole scene. One of the axioms of street lighting research is that in order to be able to see something, there must be a contrast between the target and the background (their luminances, their colour or both); when the contrast is subdued from the original, targets are seen later, slower and with lesser probability. The three fundamental questions in visual tasks in road environments are "can it be seen", "how fast" and "what is it"[3, p. 7].

One laboratory study examined the drivers' obstacle detection distances with and without simulated rain on the windscreen, and found out that in rain, the distances were smaller [71]. This evidence confirms, at least in part, the theoretical analysis by Garg and Nayar: the obstacles were harder to detect in the rain, which can be explained by the reduction of contrast. This should not be viewed as final; one could dare to propose that it is also likely that in rain, people drive on average more cautiously, which is why they are not so dependent on the better visual performance provided for by the lighting. As a consequence, they effect of lighting is smaller than it is in good conditions.

2.5.1 Traffic safety

Although this thesis will not even attempt to evaluate the effect of specific lighting attributes on traffic safety, such research is one of the intended use cases for the feedback device presented in later chapters, and it is therefore reasonable to give an overview of how lighting can affect the traffic safety, i.e. the possible models of attention and cognitive processing through which the actions of road users may lead to lower numbers of accidents.

Road traffic safety has been under heavy scientific interest in part due to the annual costs of accidents, and advances in the accident figures have been rapid in many countries. However, as the relative risk (i.e. risk per travelled billion vehicle-km) is approaching miniscule fractions of per cent [72, p. 22], the simple and most effective environmental and technical measures have already been identified. Another consequence of the low risk levels is that the statistical analysis of accident records in different conditions or in controllably different environments is getting harder, because the surveys must last long periods to reach statistical significance. Both of these changes in the traffic system mean that research must concentrate more and more on the human factors involved. Such research will, hopefully, reveal some more prerequisites for a good lighting, one that doesn't impinge on proper allocation of driver's attention on the traffic environment.

Except for unforeseen natural phenomena or truly unexpected mechanical failures, traffic accidents always happen as a result of an action a human party has initiated. In the interest of safety the environment should then be such that all parties are least likely to take actions that could put them in a conflict situation. An accident only occurs when a chain of events is unable to restore the normal state of the traffic system.[73, p. 954] The human parties involved act based on their perception, if and when they perceive something, and whether they comprehend its meaning.

Viewing the traffic environment as a system, Fastenmeier also came to the conclusion that a chain of events leading to an accident starts with a disruption which knocks the traffic system out of the stabile state; the disturbance can come from the participants' actions or it could be an external interference. If the environment fails to rectify the disturbance, and if a change in movement is necessary to avoid the accident, but the actors involved first misreact (or fail to act) and finally fail to take evasive action, the collision is unavoidable. [73]

The way people allocate their attention is said to be hierarchical [74, p. 17] – in other words, as macro level tasks, like driving, can be divided recursively into smaller tasks, from navigation down to the physical act of steering, likewise the allocation of attention is divided amongst these subtasks, and other hindrances. Even if there is no universally accepted model of human attention available, most theories would admit that the original single channel model of human attention by Broadbent (as explained by Kramer [74, p. 3]) got something right in claiming that the higher levels of processing only receive inputs from one selected target at a time; the sensory pathway filters the signals at various stages, and the filters vary between different situations and different signals.

Combining the chain of events model and the knowledge we have on directed attention, we can draw the conclusion, that whether something relevant to the situation is above the visual threshold does not directly determine the probability of an accident. However, when too many critical observations are missed or are made too late, increasing the visibility of any one of those could well have prevented the accident.

The intermediate results of the quest for a model of attention, can, however, be utilized in assessing the outdoor lighting: human attention capacity is limited, and variable, so lighting designs should consider if and how the road users' attention can be directed to the features of greatest significance to the combined safety of all road users. In fact, road safety researchers have identified that training drivers to control their use of attentional capacity leads to a higher ability to detect and respond to potential traffic hazards in later tests [75, p. 234]. In other words, if lighting can divert the attention to the highest risk details in one place, the effect may lead to better risk assessments at the next intersection – or even on the next trip taken.

Some researchers have tried to enumerate the possible atomic traffic situations, with results ranging from a number of over three million by von Benda down to 134 when those are systematically categorized [76]. Based on the prevalence and generality of the common and accepted road lighting design guidelines, and the limited number of parameters used in those guidelines, it is reasonable to accept as given that the most significant requirements for safe street lighting are similar in different traffic situations. As a counterexample, different lighting criteria are used in different road contexts, but if lighting had an especially notable effect in some specific situation (that is when compared to other situations, not to unlit roads), that effect should have been evident by now. This would mean that it isn't necessary to test the effect of lighting separately in all 134 or millions of situations, and more general tests do produce results relevant to most foreseeable traffic situations, in the environmental context the test sites occupy. It can be noted here, that only one reviewed study (details are given in chapter 3.2.3) explored how street lighting could improve the chances of correct observations in one *specific* traffic situation: two passenger cars approaching an intersection.

2.5.2 Personal safety

As was explained in section 1.3.1, the usability of public outdoor places depend also on the perceptions of the citizens, and not on objective safety and crime statistics or professional evaluations of the risks. These can also affect how much time people actually spend outside in the dark hours.

A place that is subjectively evaluated as unsafe or dangerous can even prevent people from entering. In recently published research (e.g. [13, 44, 56, 77]), the model referred to most often is the prospect-refuge theory by Appleton, first proposed in 1975. The motivation behind the model is that people prefer environments, where they have a good sense of their surroundings (prospect) and hiding places from danger (refuge). One measure has survived from studies prior to Appleton's theory, namely ample unimpeded options for escape (escape). Later articles have argued that refuge should be divided into two, as dark spots can offer hiding places for offenders (concealment), but they can also provide protection or refuge for those exhibiting fear.[77, p. 468].

Assessing these three parameters is not as straightforward as it might first seem, as the labels of the scales do not correspond to any single adjective or physical concept, which the subjects could use as a cue [78, p. 425]. Assessing any affective descriptors of environment is never straightforward [61, p. 120]. Especially the perceived prospect depends on multiple attributes of the environment [77]. It has also been mentioned, that the lack of escape, referred to as entrapment, can be also considered a two dimensional variable with social and physical axes, i.e. the unattainability of people capable of helping, and physical elements blocking escape paths.

In Finland, the Finnish Government declared in 2008 a Government Programme to make Finland the safest country in Europe by the year 2015.[6, p. 168] This goal was restated in the Government Programme of 2011, yet the current Internal Security Programme doesn't even mention the effects of lighting on the safety outdoors. On the contrary, it only lists better road maintenance and construction choices as the means to the goal. More knowledge and a better presentation of the first results measuring the importance of lighting on the increased perceived safety, and on the health benefits of the increased night time mobility, would be one way to find cheaper and more cost effective solutions, for the benefit of increased safety.

In the study of urban design and of built environments effects on human behaviour, perceived safety has been linked to increased walking and cycling, decreased car use, people's health and well-being [60]. Based on the literature review in the next chapter 3, such effects have not been considered in detail by lighting research professionals. Likewise, on social sciences side of things, the research has not that much considered what lighting research knows of light as a psychophysiological phenomenon, but rather concentrated on the effects and inner perceptions [6]. If lighting designers and urban planners know how to subdue the feelings of danger, it can have positive effects on innovation and the social networks. Nevertheless, the connection could be explored in future.

These effects are one more reason to expect more road lighting user experiments in the future.

3 Review of user studies in road lighting

In general, the studies into the effects of street lighting on the road users' appraisals have concentrated on three effects: visual performance, perceived safety, and subjective preferences. Before the subjective appraisals were considered, the main interest was only on the real or predicted traffic safety. In recent years, also other effects have been in the center of attention, most notably the effect on mood, and on the restorative potential. "Safety by lighting" studies originated in the UK (Atkin's study [79] from 1991 is listed as the first credible study in a review by Unwin et al. [15]), where, not unlike other countries with megalopolies, some urban areas may be considered insecure, and raised illuminance levels were first viewed as a cost-effective tool to hinder crime. Prior to that the studies had focused on visual performance, and the effect on actual safety was considered a logical consequence of the increased detection distances [1, 5]. Some published statistical dependence figures validating that lighting was effective in reducing traffic accidents, were established between 1959 and 1984 in road lighting recommendations [4, 5].

In this section the reader will find the studies into the road users' opinions, preferences and assessments summarized. Some visual performance studies are also included to describe the test setups and factors that have been considered, or should have been. Starting from chapter 3.11, a summary of good practice and things to consider when planning future tests is presented. The summary and list of considerations is based on the reports of the referenced studies.

3.1 Studies advancing research methodology

Light source technology and the traffic environment have advanced considerably from the gas lanterns used as the first street lamps, both in lumen output, efficiency and light distribution. Along with the changes, and with the knowledge accrued so far, there have been some points when new research methods were needed. The original studies' results might not have any practical value as such, but the methods used in them, and the challenges identified with their application can help to avoid future challenges or even mistakes.

3.1.1 Experiment designs and initial illuminance recommendation, de Boer 1967

De Boer's work on determining average recognition distances and lighting acceptability on roads with different luminances is cited in many later studies. Although published in a comprehensive book [1] on outdoor lighting design in 1967, the tests had been performed in the preceding decade. These tests included Landolt ring orientation tests, outdoors object detection tests and recording subjective appraisals.

In the first reported study [1, p. 21], based on their figure, the Landolt ring orientation tests contained about 3800 samples distributed between 0.4 and 10 cd/m^2. They then determined the required luminance level for any detection distance from that data, but without reasoning De Boer used a threshold of 80 %

correct answers as the sufficient level. Because the task of identifying the direction of a gap in a symbol has little resemblance to real visual tasks necessary in the traffic environment, they switched to using 20 by 20 cm square targets acting as objects of sufficient size to be capable of causing danger to the driver.

De Boer did acknowledge that the visual task chosen is arbitrary, and can not be the final and best task for which to design street lighting, but rather reasons why the achieved results are a reasonable baseline and that the requirements derived in his study can not be over-strict. In the end, he called for further experimental studies.

De Boer's text included a description of detection distance test[1, p. 24] with moving subjects, but failed to tell us anything which could be used to assess the reliability of the recorded distances. Test subjects in a car moving at 50 km/h indicated when they saw the test objects similar to the ones described in previous paragraph, but even at that speed, measurement errors and variability can affect the results. The subjects only numbered four, and they all repeated the test on 24 roads, where each road had two test targets.

Combining these results and previous studies, de Boer stated that 2.2 cd/m^2 is required for detecting such objects in time in normal traffic conditions.

Finally, De Boer reported on recorded appraisals of 70 different streets' lighting by "qualified observers" on a 9 point scale from "bad" to "excellent".[1, p. 27] This appraisal scale has since then been used in majority of the studies, and is even referred to as the "De Boer scale". The qualification of observers referred to the subjects being Philips Lighting Laboratory professionals or lighting engineers of large Dutch towns. Reflecting on the range bias effect described in chapter 1.1 and on the effect of professional knowledge, the relevance of this data to practical lighting design today can be said to be purely indicative. The data does support the very general notion that at typical street lighting scenes higher luminances typically lead to better appraisals.

Reported later in the book, the same appraisal scale was used to assess different lighting conditions and luminance uniformity settings (in a simulated road view, consisting of a miniature scale road model built in a 7 meter long tube with moving apertures imitating forward movement - the subjects got the impression of a 350 meters long vista). Again, with only 64 male observers and a preset level range, the appraisal values aren't important, but the correlations in the data are obvious: more light, and more uniform lighting give higher appraisals.[1, p. 66]

De Boer did, however, conduct have a behavioral study [1, p. 28] to assess how drivers perceive and act based on the luminance level present on a road. The road environment and headlight regulations have changed in this regard so that the numbers obtained don't seem to be relevant anymore, but the study setup might be adapted to the present day. In that study, they observed cars on low traffic level roads, and counted cars with their headlights off, on dipped beams or on high beams. The third category was cars with only their side lights on, i.e. drivers who considered that they didn't need more light on the road, but needed signal lights for other road users to notice them. It should be noted here that the then prevailing dipped beam lights were considered more glaring than present day technology, and less capable of lighting the road ahead.

When the proportions of those categories were plotted against the average road luminance, they could say that when the luminance drops to below 1 cd/m^2, 80 % of drivers already consider the lighting insufficient and turn on their headlights, but already at 15 cd/m^2 half of the drivers passing by had turned on their side lights. This would indicate that they had had difficulting observing other vehicles. They used low traffic roads in the study so that drivers' decisions whether to turn on headlights would not be influenced by other cars. Even in their darkest conditions of 0,04 cd/m^2 only 10 percent of drivers used high beams, which can only mean that the roads examined weren't that low traffic after all, or that the roads were in areas where many drivers wouldn't use high beams in any case. In the end, De Boer recommended that for certain the road surface luminance should exceed a minimum of 2 cd/m^2.

3.1.2 Pedestrians' visual needs, Caminada and van Bommel, 1980

Their study is referenced in several scientific papers, but was not available for inspection within this review. From what later sources tell us, Caminada's and van Bommel's test setup can't be reconstructed, but some findings relate to later user studies. They proposed based on 20 observer's results that the visual needs of pedestrians include facial recognition (identity and intent), obstacle detection, visual orientation, pleasantness and comfort.

In their results the visual comfort was to a great degree determined by the perceived glare. They found a dependence between illuminance levels and identity recognition distances, and that a sense of security could be ensured if the illuminance established a minimum facial detection distance of at least 4 meters. Hall's theory on zones of close and far spaces was given as a plausible explanation for this distance [80]. They had measurements of vertical illuminance, horizontal illuminance and semicylindrical illuminance, and out of those they viewed semicylindrical illuminance as the best measure for the detection of others [81, p. 39].

3.1.3 Test method to raise the processing level of test stimulus, Akashi et al. 2007

Although the main objective of the reported study were the effects of different spectral power distributions from different luminaires on the speed of response to peripheral vision moving targets in mesopic conditions, the test setup is noteworthy. Respondents drove a car back and forth on the same streets, where the street lighting luminaires were alternated between runs, and while looking at a fixation target, they reacted to a flip dot display some distance away from the road. The display simulated an object moving towards the road or away from it.

The test subjects were forced to utilize their higher-order decision making when reacting to the artificial stimulus, either by braking or accelerating depending on the direction of the apparent movement. At the lowest illuminance levels the difference between the response categories increased, i.e. it took even significantly longer to process the stimulus when the appropriate reaction was to accelerate, whereas braking responses were only somewhat slower. This difference existed both in relative and in absolute time differences. The authors reported this as an indication of higher decision level processing intertwined with the lower neural processing levels.[82]

3.2 Studies on visual performance

When de Boer had laid out the reasonable minimum luminance, more tests followed as new light sources were introduced, and the results were occasionally used to define and refine the national road lighting standards. Although not mentioned in his works, the gap between scotopic and photopic was known to be an understudied element in the visual performance. In the 1990's interest in mesopic research was gaining interest, and a serious effort towards an internationally agreed performance based mesopic model was launched in 2002.[83, p. 18] After that, there have been new studies comparing the previous knowledge with the newly gained insight to the mesopic range.

3.2.1 Mesopic and photopic detection threshold in foveal and peripheral vision, Eloholma 2005

Within her dissertation [83, pp. 13–17], while developing the two mesopic models, the practical linear model and the more complex chromatic model, later adopted by the CIE, Eloholma had conducted visual performance tests with pedestrians. These user tests were done in the underground tunnel in Helsinki University of Technology, with a half scale road model. The stationary test subjects indicated when the moving target pedestrian became visible, or disappeared from their view into the darkness beyond the last luminaire, and in one test series they indicated when the target pedestrian was waving his hands. These tests were repeated with various light sources and dimming levels, and with the test subjects alternatively looking at the target, and sometimes looking at a fixating area close to them forcing the target pedestrian to appear in their peripheral vision.

The results of the experiments were, that even in mesopic luminance levels when the visual task happens completely in the narrow foveal area, the photopic $V(\lambda)$ spectral luminous efficiency function is sufficient in determining the expected visual performance, i.e. how small relative contrast is required for a target to be visible against the background. However, within the mid-mesopic range (luminances of 0.1 to 0.5 cd/m^2) when the target object is outside the foveal area, the spectral sensitivity is shifted towards shorter wavelengths. Eloholma noted that this achromatic detection task is one of the fundamental tasks in road environments, and the shortcomings of using the photopic $V(\lambda)$ in mesopic conditions should be studied with more tasks. The street lighting norms could be re-evaluated if such tasks are identified, which generally occupy the peripheral vision, i.e. those tasks for which the $V_{mes}(\lambda)$ mesopic luminous efficiency function is a better predictor of visual performance.

The more general and profound result was that the most significant difference between photopic and mesopic visual performance is not between assessments of static brightnesses, but rather on task dependent differences in the thresholds for detecting dynamic changes between the background and target.

Although only eleven subjects took part in the tunnel tests referred to above, the results were in line with results from 123 more observers in other tests conducted within the European international MOVE (Mesopic Optimisation of Visual Efficiency) research consortium.

3.2.2 Simulated surface irregularity test, Fotios 2009

The laboratory study described by Fotios [27] used a specially constructed large box lit evenly and with a view aperture, which could be opened for a set duration. Some parts of the boxes floor could be raised while the view aperture was blocked. The test subjects had a fixation point at the back wall to concentrate on, and when the view aperture had been opened briefly, they had to report whether they saw a raised portion. The setup simulated a normally lit sidewalk with a small surface irregularity, i.e. an unexpected step one could trip on.

At the lowest illuminance levels tested, the lamp type had a noticeable effect, i.e. the spectral power distribution (SPD) – the luminance levels were balanced between the lamp types. Extensive reported data analyses back up his claim that their derived equation can predict the probability with which any subject would notice an obstacle of given height.

The relevance of Fotios' research to this thesis is the test method and that it confirms several factors currently considered key points in visual perception of outdoor night scenes. Firstly, that in mesopic luminance levels, when not viewed foveally, the detection of obstacles is related to the S/P ratio of the used lamp. Secondly, that older people (over 45 years of age in this study) perform as good as younger people in illuminance levels of 20 lux, but at 0.2 lux illuminance the required obstacle height is higher for an equally probable detection rate. Thirdly, the results support the adoption of the mesopic spectral efficiency function $V_{mes}(\lambda)$.

The article didn't try to estimate how the measured heights relate to real world obstacle heights. The visual size of the fixation point and of the detected obstacles could have been calculated and scaled to distances relevant for pedestrians' speeds, but for reliable results, such assessments would need to wait for the results from eye tracking studies that have been conducted after this reported study.

3.2.3 Visibility simulations in intersections, Rea et al. 2009

Although the study [45] did not employ any test subjects directly, the models used are based on previous user experiments used to validate the Relative Visual Performance (RVP) model.

In this experimental study Rea et al. constructed a virtual intersection with a photometrically accurate software, with simulated cars on three of the roads approaching the junction. They then used the RVP model which is claimed as a validated and reasonable measure of visibility for lighting levels found on roads. The researchers generated performance tables for various lighting conditions and observer locations. Simulations were repeated with one of the cars at 19 different locations at consecutive distances away from the intersection.

The main result was that the performance was nearly never under acceptable levels when the ambient lighting was at the illuminance level of 20 lux. This indicates that for example urban lighting beyond the road area helps road users. Secondly, the study confirmed an older finding that the visual performance reaches a minimum at the point where the object to be detected changes from a lit object on dark background to dark object on lit road. In measurable terms, the target contrast goes from positive to negative, and at some point the contrast is zero and the object can disappear from vision.

The study didn't, however, take into account the spectral distribution of the simulated lights, which has been shown to be a significant factor at mesopic levels outside the foveal vision, i.e. away from the point of eye fixation.

3.2.4 Facial characteristics recognition evaluation, Iwata et al. 2014

Various facial recognition tests have been conducted to date; some with a person identification task, others with a matching task, where subjects report whether they saw a face presented to them before or after the test. Prior studies, which are not reported here, had asked either how easy it was to recognize the identity, or had measured the distances at which the subjects reported success.

The study by Iwata et al. [84] used a different measure: the subjects were asked if the facial features, such as eyes, nose and mouth were visible, on a 7 point scale from "not visible" to "visible". They had data on 19 subjects and from six distances from the target (39 meters to 8.6 meters) under different lightings. They also assessed glare from not dazzling to unbearable. In their setup, there was only one luminaire between the observer and the target at a distance of five meters in front of the target.

Their results were a collection of smooth curves from almost "not visible" value to almost the best value of visible, for each light, between the maximum and minimum distances tested. It seems as if the results show a prime example of range bias. They also measured the illuminance at the location of the observer's eye at each assessment point. We can only guess that the reason the test distances didn't employ the full range possible, especially at the near end, was that they were more interested in the effect of glare: at the closest assessment point the glaring luminaire was already out of the subject's field of view.

Had some subjects been able to assess the visibility of facial characteristics approaching the test target even closer, the effect of range bias in other subjects' samples could have been estimated. One aspect of this report is worth remembering when interpreting future studies: their measured and reported correlation between preocular illuminance and the visibility evaluation, albeit with a high R^2 of 0.9071, is trivially explained by the distance, because of the test setup.

3.3 Studies on effects of lighting on attention

3.3.1 Restoration and directed attention, Nikunen 2012

As a part of her PhD. dissertation, Nikunen conducted several tests with the main focus on the attention restoration theory (ART) and how lighting can affect the restorative qualities of the night time environment. The premise of ART is, that the capacity for directed attention is limited, and restoration is the counterpoint to attentional fatigue.[13, p. 20]

The first reported study assessed how changing the focus of the lighting affects the perceived restoration. 35 test subjects were shown a total of six computer generated outdoor scenes in a controlled environment, where only the light distribution in the scenes was varied. The restorative capabilities were measured with the already validated Perceived Restorativeness Scale (PRS). The conclusion was, that the perceived restorativeness was higher, when the lighting was designed to give more light on the natural elements in the scene, as opposed to when there was more light on parking lots and roads. There were no significant differences between subjects with different pre-test mental fatigue levels, but it was deemed plausible, that such effects would only be visible with more extreme values of the control variables, or in real environments.[13, p. 63–68]

In the second study, the lighting was kept constant, but the street furniture had three scenarios: only urban, mixed, and greenery only. This time the 28 test subjects assessed the scenes on 18 scales, out of which 16 could be later conflated to the four components of the PRS. In short, the measured restorative capability was positively correlated to the preference, i.e. people prefer environments they perceive as restorative. However, this study did not reveal a significant (negative) correlation between fear and preference.[13, p. 68–71]

In the third study, the pictures from the first study were reassessed with more questions and with new test subjects, who numbered 41. The extra questions revealed a statistically significant correlation between PRS and preference, and they also validated the hypothesis that scenes lit to highlight the urban elements are viewed as less safe. However, only in some cases the test subjects had a significant negative correlation between perceived safety and preference.[13, p. 72–74]

The fourth study was conducted outdoors in two locations with different subjects. Total of 55 test subjects assessed the light on 10 seven-point Likert scales, four of which were the components of the PRS. Additionally, glare was rated on a scale from 0 to 3, by having the subjects walk one pole space while constantly telling their assessment to the assistant walking behind them. The study established some evidence that perceived restorativeness is linked to pleasant lighting, and that both brightness and dimness can evoke positive and negative experiences. [13, p. 75–84]

In the last study in the dissertation subjects filled in seven Likert scales outdoors, and a linear regression analysis revealed that the perceived colour quality is strongly connected to both perceived safety and pleasant lighting; other assessed qualities (brightness, evenness, extensiveness, lack of glare) had weaker connections, but can also be influential attributes. That is to say, that the interaction of said attributes with the perceived safety and pleasantness is not trivial; the author suggested that areas with crime problems or very low illuminance levels could yield different results.[13, p. 85–88]

3.3.2 Identifying critical tasks for pedestrians, Fotios et al. 2014

Prior studies had identified that people don't necessarily see things even when their eyes fixate on said object, i.e. their attention was elsewhere. Eye tracking studies had analyzed how people spend their time looking at various parts of the environment, e.g. when walking, but those had not revealed which gazes are critical for the safe movement of pedestrians or for their other goals.

Fotios et al. reported their study[85] where they had the test subjects press a button in response to an auditory signal (repeated at random intervals) as fast as possible. When the reaction time was delayed from their baseline, they deduced that the subject's attention was on something more critical ("critical fixation"), which delayed the reaction to the test signal. With concurrent eye tracking data, they were able to identify that the proportions of critical fixations on persons and on vehicles were higher than their proportions of all fixations; the change wasn't statistically significant, most likely because for this part they only considered data from ten test subjects, but the result was taken as a validation of the method. The authors concluded that those categories are the ones posing a potential threat, if their future path is predicted erroneously.

Then, with more participants they were able deduce that the number of critical fixations was a more robust measure than the plain number of fixations on persons, which suffers from stimulus bias. The practical and statistically significant results were that people look at other people more when they are still far away (far being distances of over 4 meters in this study), but they look at the path mostly in their near region. This seems logical, since identifying possible trip hazards is more relevant when they are near. When other people are still further away, assessing their intentions and identity still leaves time for an appropriate response, whereas visually small actions by others need no response in the near region.

3.4 Studies on perceived safety

In the 1960's, the national lighting standards didn't refer to perceptions of safety[1], but a reason to install road lighting was the reasoned effect on actual traffic safety. By the 1980's it was acknowledged that even unfounded perceptions of danger can for example prevent residents from going out in the dark and generate stress.[79, p. 1] Since the comfortability of street lighting was already known to vary between installations, the obvious question was, whether the lighting could affect the perceptions of safety?

3.4.1 Crime database review and a survey in relit area, Atkins 1991

Already in the introduction of Atkins' study [79] he recognizes that large scale attitude surveys are hampered with practical difficulties in gathering the data, or with insufficient data for reliable conclusions. At that time, they had workers run around with papers and pencils to collect the answers. They also note, that "good lighting" isn't unambiguously defined, so changes to improve lighting are different in all areas. In fact, they didn't even attempt to describe the light installation in any way [15], prior to the changes, nor after. We can only assume the upgraded lighting was "better" by the lighting standards of that day.

The first part consisted of comparisons between a database of reported crimes, and changes to their frequency after the lighting in that area was improved. In the second part they used questionnaires to assess the residents' opinions in one of those areas, and a nearby control area, seven weeks before and after the new lights were installed. The total number of completed questionnaires was 922. The question sheets had different scales in the questions: 4, 5, 6 or 9 point scales were used (and an additional "don't know" option). Out of the 26 scalar questions, most of them were answered on a continuum from very safe/very unsafe, very worried/don't think about it, very satisfied/very dissatisfied, and agree/disagree. or similar adjectives. Numerous other questions only had yes/no answer options. Results were given as regression equations for each question. As a significant outcome, they stated that there was a statistically significant increase in perceived safety, but in women only.

3.4.2 Before/after study of three locations, Painter 1996

The study[17] reported by Painter in 1996 was conducted between 1988 and 1991. The research followed the effects of improving the street lighting in three separate areas, two of which were described as deprived and which were selected by a team of crime and lighting professionals. The implied expectation of the researchers is thus that the initial perceived safety was poor.

The exact questions used on the questionnaire forms were not reported, but the ones that are given, indicate that the questions only had yes/no answers as the options. Likewise, the exact details of the improvements are not known, but the new installations met the standards used at that time, whereas the old ones did not. Old LPS light sources were replaced with brighter HPS lamps.

The research contained extensive pedestrian surveys before and after the refurbishing, with a grand total of 1046 completed responses. 12 months later, some locations included a follow-up survey to assess the long term validity. Even if 70 % of the respondents in the after condition had noticed the lighting changes, there was a significant increase in perceived safety also among those who had not noticed the improvements. Other questionnaire results were, that the respondents' fear of several individual crime types had decreased, and the results 12 months later confirmed that the residents perceived that crimes had become less common in that time period. Beyond the text, the report only contains a number of listings of percentage changes in the study variables.

As an objective measure, the number of crime reports from the areas (6 weeks prior and after the relighting) dropped across all test sites. Traffic counting also confirmed that the amount of pedestrians walking on those streets at dark hours increased in all locations, from 34 % to 101 %; even their walking pace and location on the street changed.

Later metastudies (e.g. [15]) have questioned the significance of some of these findings, as Painter's report didn't describe formal statistical analysis of the observed changes.

3.4.3 Four assessment studies of parking lots, Boyce 2000

Boyce et al. conducted four different tests in 1995 and 1996 to assess how the amount of light and other parameters affected the perceived safety on several streets and parking lots, and preferences of the test subjects. The final results of all studies were then reported [18] together in 2000. In the first studies, the respondents filled in questionnaires at each test site, with semantic differential scales including general goodness, brightness, evenness, comfortability, glare, extensiveness of area, matching to the site and riskiness of walking at the site at night.

In the first studies, all appraisals except for glare seemed to follow each other. Also, the answers followed the measured illuminance levels, and because the different illuminance measures (horizontal, vertical, semicylindrical) had very strong correlations between each other, the horizontal illuminance was used in the calculations as it is generally used as the design criteria. On closer inspection, the answers fit a hyperbolic curve when plotted as function of the mean illuminance. For a mean rating indicating that lighting was a good example of security lighting, the illuminance required was 40 lux. There was also a gender difference, 35 lx for men and 60 lx for women.

The third study compared parking lots at day and night to the street results of the previous studies. Their conclusion was, that with sufficient lighting, walking in a parking lot at night can feel almost as safe as it was in daytime, but only compared to similar locations, i.e. in general at night an average urban parking lot can not feel as safe as an average suburban parking lot.

In the last study, the subjects assessed three differently lit large bays at one parking area, both as such and with dimming glasses used to simulate a lower luminance. Some questions were substituted with rating tasks about the appearance people under each lighting, perceived safety of leaving one's car parked there and the general safety of the lighting. The respondents also ranked the three bays for two safety questions, for identification of approaching people and under which lighting people looked the best.

In the last study, the lamp type and the simulated dimming had a significant effect on the perceptions of safety, appearance of people and comfort of the lighting. The amount of light had the dominant effect on the appraisals, but at the lower end of the brightness levels used, the other aspects of the environment and of the lighting began to have a bigger impact; the variance in the appraisals increased as illuminance decreased.

3.4.4 Locations assessed for several attributes, Blöbaum & Hünecke 2005

The article [77] from 2005 describes a study exploring the effect of personality differences on the perception of public spaces after dark. The researchers had selected eight locations from an urban campus site, with different lighting, physical characteristics and theoretically motivated cues of perceived danger. Locations were selected in a preliminary study where participants had rated even more places on each attribute. Each location had different perceived prospect/concealment, escape/entrapment, and lighting.

Total 122 students, aged 18 to 32, answered a questionnaire at each location. In the end, a final questionnaire measured personality traits and background variables.

Their data revealed a statistically significant difference in their perceived personal safety between men and women, but further analysis showed that the effect depended on psychological gender, i.e. on femininity, but not on masculinity. They speculate that the reason they didn't see any effect of age was due to the constrained age distribution. All three physical cues of the places had a significant effect.

Only the interaction between gender and entrapment could be confirmed. In analysing the effects between those cues, they note that concealment had a higher effect under high conditions of entrapment, and a lower effect under low conditions of entrapment – Blöbaum and Hünecke question their evaluation on the basis of the relatively high standard deviations in their data, but a similar result has been repeated in Boomsma et al., described later in chapter 3.4.15 of this thesis.

In the end, they were able to evaluate the significance of each variable to the perceived safety. The entrapment was the greatest predictor, and biological sex the second; the next places were occupied by concealment and goodness of lighting.

3.4.5 Perception differences between different SPD, Rea et al. 2009a

Rea et al. reported in [86] of five consecutive tests, where they compared lamps of different SPD. The light sources used were MH and HPS, dimmed to different illuminances with a neutral filter fabric.

In three of the five tests the respondents evaluated, by a forced choice, the perceived brightness, suitability for social interaction, and perceived safety. In some of the studies the participants could stand between the differently lit sections and look back and forth as they pleased, in others they had to remember the setup previously looked at. Unsurprisingly, higher illuminance levels were deemed safer, better and more suitable, and only within matched illuminances, the light with a better colour rendering was considered safer.

In the last two experiments, the focus was on differences in facial recognition with the same lamps. Here, the colour rendering of the light did not have any effect, when the vertical illuminance on the face to be recognized was identical.

However, their final statistical analysis implies that the MH light sources might in fact provide steeper increments to perceived security, than an increase in illuminance would. This is attributed to the fact that at the same illuminance, the MH light provides also a higher perceived brightness. Rea et al. even contemplate that perceived brightness is a more reliable psychological criterion than the perceived security.

The results do provide an indication that subjective measurements of colour rendering or of the SPD can be indicative of the perceived safety.

3.4.6 Perception differences after a change of SPD, Knight 2010

Knight's report[21] details the study conducted in various countries. The study setup used various methods to overcome for example the possible biases in test subjects' answers, and because of that it is explained here in more detail than most of the other studies.

At the test sites, the CCT was raised and light source changed. Knight acknowledges that likely also the light distribution was altered, because the luminaire reflectors weren't changed with the lamps; however, the measured average illuminances were reasonably similar.

Assessed variables by 356 persons were perceived comfort of the area (being at ease – uneasy), quality of lighting, safety and brightness in the before and after conditions. After the switch of lamps, assessments included the perception of the change of pleasantness, safety and brightness. In some tests a group of subjects took only the first questionnaire only after the lamps had been replaced, to verify that the mean values of the assessments weren't influenced by the first round of questions.

At identical illuminance levels the perceived safety, perceived brightness and comfort were higher under MH lighting than under HPS lighting. Also facial detection was possible from further away.

In the UK location they had comparison samples: some neighbouring streets had a lamp type change in another direction, so that Knight could rule out a general tendency to perceive changes after any change of lighting. The residents participating in the evaluations weren't told that the focus was lighting, but perception of safety after dark.

Volunteers visiting the areas both before and after the change participated also in a face recognition test. The faces were pictures of local celebrities on paper, four faces each time, counterbalanced between test subjects. Subjects walked toward the photos from a distance of 15 meters until they were close enough for the three stages of identification (gender, guess of and certain of their identity). No reasoning was given for the selected initial distance, and most subjects identified the gender already from their starting point. The reported values are in the same ballpark with prior studies, which were omitted from this thesis.

The report identified the streets in St. Helens, UK by their name, but the locations in the other countries were only referred to by the name of the city, and can not be revisited, not virtually nor in person. However, it turns out that the streets in St. Helens were in two different locations roughly 6 km's away, which is in contradiction to the clause "in neighbouring streets" used in the paper [21, p. 316]. Also, one of the locations can't be identified accurately because currently, Wedge Avenue does not ever meet The Shires, the gap being approximately 6 km – unless the test subject group was in fact divided to two locations. Knight reported that in Hereford Close, against expectations the facial detection distances were longer under HPS lighting, whereas MH light sources provided longer distances in all other places (that street was the comparison sample, where lighting was changed from MH to HPS). From the pictures taken at Hereford Close and other streets three years later, one can say that said street's geometry is different from the other streets included in the study: the longest straight section of Hereford Close is just barely long enough to accommodate two street light poles, whereas the other streets include longer straight parts.

Knight lists some possible factors for the unexpected variations in the facial detection distance: glare, light distribution, lamp spectrum, vertical or semi-cylindrical illuminance, but doesn't mention that the last one, semi-cylindrical illuminance, has been mentioned frequently in other reports of facial recognition tests, and in relation to the perception of the three dimensional forms[87].

3.4.7 Comparing different demographic groups' appraisals, Johansson et al. 2010

Johansson, Rosén and Küller had the hypothesis that various demographic groups give different mean values for some assessed subjective measurements of the outdoor lighting, and that there is a relationship between personal environmental trust, perceived visual accessibility and perceived danger.[88] This had some basis in previous studies, e.g. in the work by Blöbaum & Hünecke[77], explained in section 3.4.4. 81 volunteers from three demographic groups, the visually impaired, elderly people and young women, walked a 170 meter route in a park-like environment.

They had 24 questions: five from the work of Blöbaum & Hünecke, 15 items for lighting quality and four for the environmental trust, all on five point Likert scales. The weather was similar on all nights. Few visually impaired volunteers had to have an assistant walking some meters behind them.

In the analysis, the separately measured lighting parameters were so tightly correlated with each other, that they used only two of them, cylindrical illuminance at ground level and the difference in horizontal illuminance from the path to the grass - basically a measure equal to the surround ratio. They were able to use hierarchical regression, because, as they report, samples of more than 60 subjects are acceptable for the method. For some analyses, several items were combined to form indices of e.g. hedonic tone, a term which has been used previously and roughly equals pleasantness. In the referred study, hedonic tone was the mean value of unpleasantness, unnaturalness and monotony of the lighting. 45 % of the variance of perceived danger could be explained by the assessed scales (not only hedonic tone).

As a result, the personality trait of environmental trust was confirmed as a factor mediating the perceived danger. Also, the perceived visual accessibility was linked to that trait.

3.4.8 Exploring the effect of semantic priming on appraisals, Unwin et al. 2010 & 2014

The papers where this method was described also contained assessments of past studies into the interaction between the perceived safety and the lighting. That metastudy is described later in chapter 3.7.1.

In the latter part of their article [15] they describe their new test method. Their subjects provided two photos, one where they would and one where they would not like to walk alone after dark. After that, the subjects were interviewed, asked to describe properties of roads they would feel safe to walk alone in the dark, and after that, their photos were available to them when they were asked to describe why they had chosen those specific places. Finally, they were shown four photographs taken by the study conductors, and asked the same questions; the photographs included only two places, with different apparent illuminance levels within the picture pairs.

The results of the described pilot study were used to tune the methods, and later the same authors reported the final results with 53 participants [89]. In short, the results show that both with and without the semantic priming the most often reported features of places where people would confidently walk alone, are access to help, and presence of lighting.

3.4.9 Luminaire comparisons on pedestrian ways, Jaatinen 2010

As a part of her master's thesis [90] Jaatinen compared pedestrians' appraisals on two roughly similar footpaths lit with LED and MH, their largest difference being a different vertical slope. Test subjects were divided in two groups so that the sets of answers could be compared with each other. The first group repeated the test on both footpaths in autumn, and redid the test on the LED-lit path again in winter. The other group only tested the latter path in winter. At the end of the path the respondents filled in a questionnaire, where some answers were simple yes–no values and others were semantic differential scales.

Because the respondents repeated the tests, hence knowing the questions while walking, the form included mock questions about other aspects of the environment not related to the lighting; the objective was to guide their focus of attention away from the light and away from the test objects for perception. The real questions asked about the perceived safety, light distribution, glare, amount of light and the pleasantness of the colour of the light. Additionally, there were questions on face recognition and Landolt-C detection, measuring the rate of correct observations. As such, the actual values aren't interesting, but the differences between the tests are of value.

The concept of glare is easily misinterpreted by the layperson, so there were two different questions about it: whether the lighting was "too bright" and "caused glare"; however the mean values of the answers were nearly identical in all tests. The worst glare was perceived on the LED path in autumn. The respondents felt safer on the footpath lit with LED luminaires and the light distribution was perceived as more even. One should note that in general the safety ratings all were close to the "very safe" end of the scale. With so small samples, no statistical significance analysis was necessary.

Against expectations, in the facial recognition task the learning bias is not visible in the results, as it was in the Landolt-C answers.

3.4.10 Dynamic lighting distribution and perceived safety, Haans & de Kort 2012

The study by Haans and de Kort is noteworthy, because it is the one of the few recently published studies with moving test subjects, and the only recent study about dynamic lighting control with moving subjects. Haans & de Kort had subjects compare different dimming setups around pedestrians.[56]

First, data of pairwise forced choice between different distributions identified a effect on perceived safety. Second, subjects rated the distributions on nine questions, which were cues for prospect, escape and concealment. Again, the distribution had an effect: when the highest illuminance was concentrated near the assessor, their feeling of safety was at its highest.

Then test subjects walked with each distribution following their position. Prospect was evaluated as higher when there was more light near the test subject, and not when the higher lighting level was some distance away, as had been expected. The light distributions used were somewhat different from the ones in the first experiment, because the control method used could not send dimming commands to the luminaires fast enough to implement identical distributions in realtime.

There was no comparison to the normal lighting setup with walking subjects; this was because they had decided to use a constant light budget to suppress the effect of changing overall lighting level and with all lights lit in the regular manner the street would have been much brighter.

A possible uncontrolled effect in these tests was that the light distributions were balanced with an illuminance budget, apparently with the sum of spot measures under each luminaire. As the distance between each supporting pole was 30 meters, even with the constant illumination scheme, the area of equal luminance under each pole covered different solid angles in the observer's eye. That is to say, that especially in the ascending lux level setup (more light away from the observer) the overall perceived lighting level was not equal to other setups, even if the sum of luminous output of all the luminaires was the same. This could be argued with knowledge that adaptation luminance of the human eye is determined from a space angle much smaller than would be required for the illuminance budget used in this to produce a constant adaptation luminance. For the space angle relevant to adaptation, literature [25] cites an angle of 20 degrees but recent research, e.g. by Cengiz [91] and by Uchida et al. [92], is still exploring angles smaller than that, at least for mesopic conditions. The adaptation level by itself has not been considered in studies of perceived safety.

Haans & de Kort proposed that the preference stems from the concepts of action space and vista space, which divide the world into a near area where we can operate, and outside area, in which actions are only planned. The usually considered ranges for those concepts happened to coincide with the results of this study. They called for more research, as they already identified uncontrolled variables and places for technical improvements in their setup.

3.4.11 Luminaire comparisons on pedestrian ways, Rantakallio 2011

In a chapter of his master's thesis [93] Rantakallio examined some of the then recent LED luminaire installations in Espoo, Finland with users. The user studies included a face recognition distance task and a questionnaire form with nine questions with a five point scale (surface illuminance uniformity was rated on a three point scale).

Even though the questions included perceived safety, there was no detectable dependence between the safety perception on the measured luminance values or the other scales. The other questions are analyzed later in chapter 3.5.4.

3.4.12 More luminaire comparisons on pedestrian ways, Rantakallio et al. 2012

A joint research project of Aalto University and several cities, used in part the user tests Rantakallio had come up with in his master's thesis, reported above. They did add and replace some questions, and they also had a glare measurement task where an assistant walked behind the test subject as they spoke out their continuous assessment of glare. This was the same task as in Nikunen's dissertation, from chapter 3.3.1. However, in the project's final report [20] they listed all the questions, their means and standard deviations and the answers to open ended questions.

In their analysis the most important factors for a comfortable lighting in pedestrian environments were suitability to that specific area, pleasant colour reproduction and a good sense of security. Suitability for dark time leisure walks was mostly based on the aspects related to the attention restoration theory, the comfortability of lighting and suitability to the area.

3.4.13 Changes in appraisals after retrofitting, Kuhn et al. 2012

The study in a residential area in Sweden, by Kuhn et al. [14], asked residents to assess the perceived safety of the area under study, among several other subjective attributes. However, the initial level of perceived danger was very low, and there was no change in perceived danger when the lighting was retrofitted. Because of this, the whole study was more related to studies of other subjective attributes, and is explained in chapter 3.5.5. However, this reminds us that while reading the studies on the matter at hand, it is good to keep in mind that sometimes the public does not consider the places under scrutiny as dangerous, even after dark.

3.4.14 Discomfort glare ratings, Lai et al. 2014

The established equations for assessing discomfort glare assume that the glare source's luminance is constant across the light emitting surface. This has not limited the applicability in road lighting, because the light emitting surfaces used to be relatively large and even, but as LEDs gain popularity, the luminaires consist of several small light sources, each with a very high luminance. Lai et al. had 10 subjects (from 25 to 40 years of age) stare at a simulated road environment (projected image in a dark laboratory), with reference and test lights appearing within the picture for 3 seconds, with repetitions. The mean values of their glare assessments were only reported in a figure, but estimated from that they varied only between 3 and 7, on that 9 point scale. [8]

Their data revealed that to perceive a significant difference in glare, the luminance of the array light source had to be changed less than that of the reference.

Also, when the reference luminance level was higher, the difference between uniform and nonuniform surface luminance was smaller.

They then had users compare the comfortability of the sources, which were adjusted to provide identical illuminance at the observer's eye: the array light source was less comfortable up to the distance of 20 meters. After that distance, the subjects' comments revealed they didn't see the individual LEDs anymore so the experience was identical.

As a conclusion, they state that this can be a relevant finding for pedestrian environments, but not for drivers; other than a driver of a convertible car looking straight at nearest lamp, the distance to the luminaire is high enough that the array of LEDs can not be distinguished. Also, as major roads employ luminaires with higher luminance, the effect is even less important to drivers.

Although it was first necessary to test if the effect which they found exists, their study didn't address the fact that they only used one placement of test luminaire in the visual field, near or at the fixation point, both in the laboratory study and in the outdoor study, yet they derived conclusions about glare sources in the periphery. In other visual tests the results have sometimes been a function of the location of the test signal in the visual field, and this would be a trivial next step to explore.

3.4.15 Entrapment, lighting and gender, Boomsma & Steg 2014

Boomsma and Steg wanted to develop an understanding of how reduction in lighting affects the public acceptability of the lighting policies delivering and defining those reductions. They justified with previous studies that if lighting changes decrease the perceived social safety, the new lighting levels will not be acceptable. By validating the findings of previous studies, they wanted to show which physical characteristics of the environment and which personal characteristics are relevant to the assessments of perceived social safety.[44]

The 88 test subjects, first year students, were shown four computer generated videos of 40 seconds each, simulating a walk in a built up area, on a computer display in a dark room, and they were asked to visualise themselves as walking in that area. The variations had the four combinations of high and low entrapment with high and low lighting levels. Entrapment was modelled by altering the width of the urban street from 20 meters 5 meters. After each video the test subjects gave their ratings for perceived social safety (with 7 questions on a 5 point Likert scale), and rating of acceptability of the depicted lighting on 5 scales. In analysis of the data the last item of acceptability, assessing the amount of lighting, was considered a separate indicator.

Against their hypotheses, there was no difference in perceived safety between high and low entrapment settings when the light level was low, but rather when the light level was high there was such a difference. This was, however, in line with the findings of Blöbaum & Hünecke [77], in which, contrary to their expectations, men and women didn't give different assessments of acceptability. In Boomsma et al.'s data, there was a noticeable difference in the perceived safety between men and women when entrapment was varied: men didn't consider either one as less safe, but women preferred the lower entrapment setting.

They also reported that in low entrapment settings increased illuminance increases the acceptability more than the same lighting change did in high entrapment setting. Blöbaum & Hunecke had reported on a similar effect, but on perceived safety, yet Boomsma & Steg couldn't replicate that finding here. This could be explained by the fact that adding more light doesn't normally open any new escape opportunities, whereas in a low entrapment environment the number of easily visible escape options seen can increase.

The authors did admit that there could have been an intermediate lighting level variant in the test, which would have conformed to the applicable local lighting norm of 15 lux. The simulated videos were reported to have portrayed lighting levels of approximately 12 lux and 17 lux, as assessed by two light experts, but no details were given as to the assessment method. We can only take it for granted that the monitor was suitable and the assessment based on appropriate and sufficient measurements. Boomsma and Steg also noted that the gender distribution of respondents wasn't equal, and likewise, the respondents were relatively young. They proposed other age groups should be studied before generalizing the results to all ages.

3.5 Studies on other subjective attributes

There's more to subjective appraisals than comparisons of installations or rating scales of perceived attributes or the strength of affective responses. Subjective appraisals also include the motivations and the perceived value of light, and of its attributes, to the difficulty of everyday tasks, and to the experienced emotions. In the end people can value things differently in different contexts, so the background variables possibly affecting the results are numerous, and the results of the studies best read with that in mind.

3.5.1 Spatial perception differences inside a car, Caberletti et al. 2007

Although the study [94] was concerned with the effects of vehicle interior lighting on the driver's perception (as opposed to street lightings effect), it still has noteworthy findings. They compared the same car with different interior lighting setups in a driving simulator depicting a driving task on a lit highway. Each variation was rated with 18 semantic differential pairs concerning space perception, perceived interior quality, interior attractiveness, perceived safety, alertness and functionality. They used a continuous scale, i.e. the subjects would draw a mark freely between the scale extents.

The interior lighting mean luminance levels were limited to between 0.007 cd/m^2 and 0.1 cd/m^2, leaving it below the luminance levels of 0.1 to 1.5 cd/m^2 encountered on the simulated driving lanes. The results revealed that especially the colour of the interior light influenced the space perception and the perceived safety, but also other parameters did have an effect on the other assessed characteristic. However, there was no sign of any of them influencing the emotional state of the test subjects.

Mean values of the answers showed, that a small amount of light is considered better than no light in the interior, but long before luminance reached the driving lane luminance, it was uncomfortable and distractive. A statistically nonsignificant effect of the bluer light was a better perception of the car interior, which was also accomplished by adding light on the doors, which appear on the peripheral vision.

The relevance to this review is the dependence of perceived safety on the colour temperature in the peripheral vision, at least in such tested road conditions with few visual distractions. This was one of the few studies encountered with a continuous appraisal scale.

3.5.2 Discomfort glare ratings, Bullough et al. 2008

Published in 2008, their article [95] described a set of studies conducted to build a model for predicting discomfort glare from outdoor luminaires; some of those indoors, some outdoors. They controlled the surfaces' luminances, light source sizes, viewing distances and secondary light sources visible in the indoor tests, and transferred their gained knowledge to non-uniform backgrounds outdoors.

Their main findings were, that the subjective ratings of discomfort glare were influenced most by the illuminance at the cornea produced by the glaring light source, at least more than by the luminance of that light source. Their model with three illuminance parameters (light source, light source surround and ambient illuminance) already had a goodness-of-fit of r^2=0.70 to the combined data of all their experiments. However, there is a possibility of inherent self-correlation when the equation is in part based on some of those numbers.

The relevance of the study to this thesis is, that it would be – in theory – possible to predict the discomfort glare at any location with the three parameters measured with an imaging luminance meter, but that the predicted assessments would still contain uncertainty.

3.5.3 Preference between MH and HPS, Ekrias 2009

Most of Ekrias's PhD. dissertation's studies were photometric in their nature, but in the seventh article he presented a study with test subjects finally selecting their preferred lighting setup out of the two setups used [96]. Apparently subjects also answered open ended questions as to why they preferred their choice. Before that, they had repeatedly compared the visibility of two targets, with a five point modified Likert scale for each pair of targets.

The study only reported the result (most subjects, 67 %, preferred the MH lighting over the HPS lighting, and 17 % deemed them equal), and that the reason to like MH lighting was that the colour temperature of the light was perceived as more natural and pleasant. The few that liked the HPS light better, thought that the light was warmer and that it seemed more familiar.[96, p. 84].

In view of other studies' results, the preference for warmer and more familiar light could be an example of the evolutionary roots, where low colour temperature light generated with fire meant safety at night[13, p. 94–95].

3.5.4 Subjective appraisals of installations, Rantakallio 2011 revisited

The master's thesis of Rantakallio [93] was explained cursorily in the prior chapter 3.4.11 for its relevance to studies on perceived safety, but it also contained users' appraisals of other subjective attributes. Rated items were whether there was sufficient light, the brightness, uniformity, glare, luminaire brightness, light colour pleasantness, colour temperature, perceived safety and general quality. In an open ended question, the test participants reported that they valued most the reliability of the light installation, sufficient luminance and light distribution evenness.

With only 14 volunteers, only a near linear dependence between the glare and the brightness of the luminaire was visible in the data. Such answers were speculated to be a direct consequence of the test subjects occasionally looking at the luminaires, instead of looking constantly at the street level like they would normally do. Already those LED luminaires proved to be a viable product for footways and cycle paths, based not only on the conventional measurable performance variables, but also on the perceived attributes. There was no preference for lower colour temperatures.

3.5.5 Preference differences between motorways and urban areas, Vikari et al. 2012

Results and method were reported [97] on a questionnaire in three parts in autumn of 2010 to those who drive a car at least occasionally. The parts were about perceived importance of light characteristics, visual tasks, and a comparison of driving on lit and unlit roads, i.e. to their perceptions and preferences, but not in the context of any identifiable road.

The responses by 105 subjects revealed, that there is a statistically significant difference between urban areas and motorways for three perceived lighting parameters: the importance of the amount of light, the discrimination of colours and the amount of light on the road surroundings.

The visual tasks studied were rated as more difficult in motorway driving, except for seeing traffic signs, which was slightly more difficult in urban areas. Out of the visual tasks, the difference was statistically significant for: seeing pedestrians, large and small animals and small objects, assessing the driving speed of other vehicles, generally seeing when the oncoming (and following) vehicle's headlamps are glaring, and assessing distances.

The last part of the study showed that, with regard to both safety and driving comfort, the lighting was considered significantly more important in urban areas than it was on motorways. Responses indicated that driving requires greater concentration on unlit roads. Also, most drivers would pay a little extra to get streets and motorways lit. Most drivers would, rather than have all lighting removed, see the amount or quality of light reduced.

There were only two statistically significant differences between age groups in their appraisal of the importance of lighting properties in urban areas: the younger valued the uniformity of light on road surface more than the older group, and the older valued the energy efficiency more. As is expected, the tasks were generally rated as more difficult by the older group.

Interestingly, the young drivers considered the lighting more important in urban areas, more than the older group, and they also appreciated their own headlamps more. Other background questions did not reveal any significant differences, i.e. the age had by far the greatest effect on the answers, even if there were some statistically significant differences between the answers to some questions by men and women.

The open ended comments received listed a good deal of different concerns and ideas. Visibility of pedestrians and other smaller sized traffic participants was one recurring concern.

3.5.6 Changes in appraisals after retrofitting, Kuhn et al. 2012

In their study the residents of two suburban Swedish residential areas were interviewed before and after a change of luminaires in the yard and footways in the vicinity of their homes. [14] The areas were enclosed yards, such that practically no light from nearby streets was present. In the first area the previous lamps were HPS and in the other area the lamps were HPM; both areas were installed with identical new LED luminaires in summer 2010 without altering the placement of luminaires, and the residents gave their appraisals the previous winter, and in the middle of the next winter. The luminaires were claimed to represent the then state-of-art best LED luminaires available.

The questionnaires were thorough, with 10 pages and a total of 21 semantic differential and Likert scales, grouped into four categories: subjective quality of light, visual accessibility, perceived danger and finally, perception of natural colour of objects in the surroundings.

The authors noted that there were two variables beyond their control that could have affected the results somewhat. Firstly, the perceived danger was rated as very low in all samples. Secondly, the other area had snow cover when the interviews prior to the change were conducted, whereas the other area didn't have any snow. The answers were however analyzed with that in mind, and the snow seemed to increase the perceived hedonic tone of the lighting. Kuhn et al. also noted that in the first area the sample was not representative of the population, but had an overrepresentation of older female residents.

Their results showed that the new lighting, despite providing the same or higher illuminance at measured spots, used considerably less energy (yearly energy use figures had a measured drop of 56 % and 76 %), but was perceived as better on most attributes, and as good as before on the rest of the scales assessed.

3.5.7 From adjective pair assessments to two quality indices, Johansson et al. 2013

Johansson et al. had test subjects evaluate the installed lighting at 10 sites, with a minimum of 22 subjects at each site. At some sites, the lighting was changed between visits. The answers were given on 16 adjective pairs on a five point Likert scale. They formed a Perceived Comfort Quality index (PCQ) from five of the adjectives (hard, warm, natural, glaring, mild and their counterparts), reasoned with factor loadings. Their index was then shown to predict the perceived danger, but also the visual accessibility. Their another index of perceived strength was not related to the perceived danger. [98]

Six of the scales used were excluded as not explaining the variance in the answers, so their final proposed study method includes 10 questions. They also suggest that these could be collected on mobile devices, and that the assessments could be tied to a location with GPS data. They note that in other cultural settings another subset of the explored adjectives might turn out as having the greatest explanation power. In their index, the scale from mild to sharp was given the highest coefficient of 0.856, but the other scales also had coefficient from 0.701 to 0.812, soft to hard being the next highest.

3.5.8 Distribution adjustments and appraisals, Viliunas et al. 2013

The study reported by Viliunas et al. [78] used adjustable luminaires on a section of a street on the university campus, next to a wood (i.e. next to darkness), with no other traffic present during trials. The research group used a lot of effort to assess and validate the adjective pairs' relevance to street lighting, first with psychology student group's mean assessments and then by analysing them on two axes.

The actual test run was similar to other studies: 21 female students answered the 12 semantic differential scales in counterbalanced order, with a range from -3 to 3, after they had viewed the seven luminance distributions selected for testing, also presented in counterbalanced order. Out of the same group, eleven subjects also assessed 15 Likert scale items, numbered from -2 to 2. Those items were designed to fit five groups, three of which related to the prospect-refuge theory, and the remaining were safety, and mask.

They then constructed unlabeled factors explaining the most of the differences in the first semantic differential scales. First factor contained pleasantness, calmness, encouragement, and niceness, contributing to well-being, and the second factor included evenness, wideness and exposure, i.e. assessments of physical properties. The conclusion was, that the only significant difference in these between lighting setups was explained by the luminous flux of the luminaire closest to the observer. The second Likert scale test revealed similar results.

Then, by excluding statistically insignificant Likert scales, there was again two explaining factors accounting for 77 % of the information, with 66 % explained by the first factor. Their report told us that the first factor consisted of assessments of prospect (it is easy to even the smallest objects) and safety twice (I feel safe on this street, and I think one could be easily assaulted on this street). The second factor was made up of two other questions assessing prospect. Again, the luminous flux of the closest luminaire caused the most significant difference to these factors between tested lighting patterns.

This study used a less pronounced difference between the luminaires' luminous fluxes than the test reported prior by Haans & de Kort [56], but the results are in line with those.

3.5.9 Novel data analysis of appraisals, Romnée 2014

Romnée and Bodart compared assessments on two streets by 41 subjects (17 to 78 years of age), and calculated satisfaction indices for several concepts identified in literature from unidentified number of questions.[99] The concepts were the feeling of safety, the feeling of visual comfort and the perception of lighting atmosphere. In their data, the newly installed LEDs were better on all these measures.

They asked subjects' answers on what factors they consider the most important for the lighting quality. Five factors with support from over 60 % of sample subjects were chosen.

The novel data analysis was that after they had calculated the indices for the concepts (some questions were included in several indices) they constructed a satisfaction score (from -100 to +100) for the five indices. Then each subject's satisfaction score was plotted on the two axes signifying the two installations. Those within a distance of under 50 "score units" from the origo were considered as "no strong opinion", but the graphs showed that all satisfaction score data points were more in favor of the LEDs, and visually inspected, a majority of scores were outside the no strong opinion circle on the graphs.

3.5.10 Use of Focus Group prior to lighting study, Kostic & Djokic 2014

The questionnaire study with 112 respondents between 20 and 30 years of age deserves a mention in this thesis, because they used the focus group method to refine the initial questionnaire to the form used with the full sample. They considered that the changes made the questions clearer and they were easier to understand in the intended way. [100]

Focus groups are also used in usability research, and this was a form of usability improvement to the developed questionnaire. In a focus group, the participants are briefed on the matter, and the researcher asks questions which the participants discuss freely. The researcher takes notes and makes the final conclusion from what was said; analysing the results of a focus group meeting is to be done with caution for a few reasons. Firstly, the opinions voiced may not represent a consensus – not all speak up to contest the opinions of others – and secondly, there is a potential for social desirability bias; sometimes people tend to answer in a way they believe others expect them to answer. A benefit of the method is that the researcher can observe how the participants defend their view.

Their report explains how there was a statistically significant difference between the mean values of each variable, but does not attempt to analyse if there were any crosscorrelations between them.

3.5.11 From users' preferences to a new luminaire, Athledics project, Juntunen et al. 2014

As a part of the AthLEDics joint project, where they examined intelligent lighting control, pedestrian acceptance and possible power savings, those involved designed a LED street luminaire which was meant to save energy and have better user appraisals from the pedestrians. Several different existing luminaires were first installed on a recreation area pedestrian road, and passers-by's appraisals recorded with questionnaires in November 2011. The questions, on a five point scale, included whether there was enough light, whether colour of the light was pleasant, if there was glare, was the verge sufficiently lit, and a overall rating on a 7 point scale.

With all the photometric data, not limited to luminous intensity distribution, and the 46 users' appraisals and comments, a new luminaire was designed. The new luminaire then replaced one of the worst performing luminaires, when a new survey was conducted in February 2013 on the same test route.

The result, with 23 new subjects and the new luminaire, was that their new luminaire outperformed – in the assessments – the other luminaires on two scales: overall appraisal and amount of light on the verge, the latter of which had been a design objective. In the rest of the questions it had the second highest appraisals, except for the colour of the light, which the users considered the worst out of the tested luminaires. The CCT of the light was higher than the older luminaires' CCT.

Later, the new luminaire was utilized with an intelligent adaptive control system, which adjusted the power of the luminaire based on the light reflected off the ground (which changed both at dusk/dawn and with snow cover). The control system also dimmed the light when five minutes had passed without any pedestrians passing by. Power savings potential was estimated to be about 28 Wh/day for each luminaire. Users' appraisals were, however, not collected with the adaptive control.

The method and results were reported in Juntunen et al. [101] and some more specific details in Tetri et al. [36].

3.6 Studies on eye gaze direction

First eye tracking studies outside lighting research were conducted over 100 years ago with electrodes on the skin measuring the activation of the muscles that turn the eyes. Modern methods received increasing attention since the 1980s. The problem with applying eye tracking research into outdoor lighting studies has been that previously, the systems were not mobile by any practical measure. Computer usability evaluations have used screen mounted eye tracking devices since the 1990s. In such settings the question of calibration is easier than in mobile, head mounted devices, which must measure the tilt and rotation of the head simultaneously. [102]

In traffic research, car mounted eye trackers have been used in the 1990s, but no references to dark time studies was found. Helmet mounted eye tracking devices were available already in 1997 when Patla et al. conducted their tests with pedestrians, but back then the subject was tethered to a computer with a 30 meter long cable [103]. In view of the above, it isn't surprising that there have been next to no eye tracking studies in lighting research prior to the study reported by Davoudian in 2012. In this review, the only older study referenced in later studies was an eye tracking test in a dim room examining what people look at when avoiding a obstacle on the ground [104], but in that the lighting conditions weren't described, other than the average illuminance.

At least in some contexts, the fixation frequency on an object is taken as indicative of greater interest on that target. A longer duration of fixation on a particular target can mean difficulty of extracting information from it, or that the target is in some way engaging [102]. In interpreting the results, it's good to remember the inattentional blindness effect, in that people will from time to time fail to notice and recognize objects even in their foveal field [16, 105, 106], but can also sometimes notice and recognize targets outside that area [107].

3.6.1 Classifying gazed-at objects, Davoudian 2012

In Davoudian et al.'s study [16], 15 test subjects (20 to 60 years of age) walked on three set routes, and their gaze direction was recorded with a head mounted eye tracking device, along with a video in the direction their head was pointed at. On the last route, they were accompanied by the researcher, and unaware that their gaze was already being recorded - they thought they were being guided to the next test route's starting point. This was to see if the presence of the researcher, which would make them feel safer, would affect the time spent looking at the pavement. In the end, the subjects briefly described the recorded sights and their experiences. Later, five of the subjects returned to walk the routes at daytime for control.

The recorded gaze directions were classified into four groups in postprocessing, with "pavement" and "other" being almost equal in the time spent looking at them, both night and day. Although subjects spent on average less time looking at the pavement at night, the difference couldn't be confirmed as significant. The remaining target classifications (people, signs, transient objects) received very little fixations. However, when scaled to the counts of them occurring in each route, at night subjects looked relatively more often at other people and at transient objects than they did at daytime, but the total duration of gazes at those features was virtually the same. The class "other" contains such features as houses, roads and trees.

The pavement class of fixations was analyzed with more detail, and subjects looked (on average) at locations about 1 second ahead of them at their walking pace, and farther. Davoudian et al. argue that the task of walking at night is in fact of low visual demand, i.e. people generally look at the pavement after they have assessed the other features of the scene. This is backed up by the finding that at night the subjects looked more at other possible hazards in the area, and that on the last route where they were accompanied, and therefore feeling safer, they again spent more time looking at the ground.

The subjects only spent about 3.5 % of the time looking at other people, but in the interviews they indicated that seeing the faces of others was considered important; yet some said to first rely on the body language of approaching people. Minor feelings of being threatened were evoked only by the behaviour of passers-by, not by their facial features, although at close ranges, where they could have evaluated the faces of others, the subjects tended to look away; facial gazes were more often allocated from farther away.

3.6.2 Gaze distributions on rural roads, Cengiz et al. 2013

Cengiz et al. fitted a passenger car with an eye tracking camera, and subjects drove a 17 km long route, with both lit and unlit sections at a posted speed limit of 80 km/h, back and forth in the hours of darkness and at daytime. After glances at the dashboard were eliminated, the distribution of gaze directions was calculated for each 25 m long interval of the route. The majority of gazes were within few degrees from the origin (straight forward) in both horizontal and vertical directions, and almost fully within -10 – 10 degrees. [91]

The size of the visual field which affects the adaptation level is still under consideration, especially in the mesopic region. Therefore, they compared the mean luminance for each interval, computed with different sized visual fields: the mean luminance of that visual field was in all cases higher in the unlit sections of the route, than it was on lit sections, but this is explained by two things. Firstly, in unlit sections the subjects spent more time looking at the surface lit by their headlights and secondly, in the unlit sections, they mostly used the high beams; at some intervals, the 1 and 5 degree mean luminances even exceeded the upper limit of mesopic range, 5 cd/m^2. When the tested visual field size increased, the variance between subjects in mean luminance values decreased.

The subjects weren't interviewed, nor did they fill any questionnaires, but the study was relevant for this thesis for two reasons. The data shows a trend that at least when driving, people mostly look at or very near the best lit area in their view, and second, because the calculated tentative adaptation levels from luminance images were similar to the values obtained from road surface luminances.

3.6.3 Gaze directions walking on pavement, Luo 2014

First reported in an article leading to her doctoral dissertation [108], Luo recorded the eye movements of five subjects walking a relatively low use pedestrian path next to a road in both directions, before filling a questionnaire. The eye tracking system used a helmet, onto which the camera recording the eyes was mounted. To complement the eye tracking data collection, participants rated 10 questions on a five point scale from -2 to 2; questions included pleasantness of the pavement, light level (too dark to too bright), light distribution uniformity, pleasantness of the colour of light, ease of facial recognition and pleasantness of walking on that street in general and alone, and their perception of the importance of lighting for both movement and safety. They were also asked to rate the possible glare, but all subjects agreed there was no glare. With only five respondents, the results of those questions are quite summarily. However, all subjects evaluated lighting as important for safety.

To analyze the eye tracking data, a picture was taken every five meters from the same middle-of-road path which the subjects had walked. Then, the gaze direction data could be matched (roughly) to the features in the relevant picture. At this point, one subject's eye tracking data was discarded as erroneous. Most of the fixations were on the pavement they were walking on, and their gaze wandered more in the vertical direction, that it did in the horizontal direction. After the view was divided into a 8 by 8 grid of cells, the analysis compared the number of fixations and time spent fixated in each cell. This showed, that some cells didn't have the highest number of fixations, even if they had the longest total duration of fixations, and vice versa. There were also differences between subjects in this regard. The problems the study conductors encountered were that the camera helmet had to be fixed quite strictly to the subjects head before calibration, which could alter their behaviour. Luo suggests that with some additions (pupil size measurement), the eye tracking system could be used to investigate the adaptation level changes in dynamic conditions.

3.7 Metastudies

Several studies assessing past research papers exist. These studies have tried to assess whether the findings from past studies are still the best knowledge that we have, or whether later studies with better or similar validity have provided contradictory evidence. Metastudies have also identified minor and major problems with the reports, i.e. many research papers do not include enough data to assess the generalizability or the statistical significance of their findings.

3.7.1 Assessment of studies' reports, Unwin et al. 2010 & 2014

The papers referenced in this chapter also contained a novel test method for assessing perceived safety. The method was described in chapter 3.4.8.

In 2010 Unwin and Fotios published their paper [15] assessing several studies, and describing their then upcoming study addressing the issues they had found in previous research. For example, they criticised the study report by Atkins et al. [79] (referenced in chapter 3.4.1) as lacking the necessary data for statistical review needed for generalizability. The other six studies they had reviewed had also used questionnaires before and after a change in lighting. Four studies used five point rating scales, one yes/no responses, and the remaining two didn't report the questions at all. The reported results had been that changes from HPS to MH or fluorescent lighting increased feelings of safety, but the number of respondents were in Unwin and Fotios' view too small for affirmative conclusions.

The paper also criticised the use of odd number of answer categories, because the presence of a middle value is then known to attract subjects away from the extremities of the scale[26, p. 410]. They hypothesised that Johansson et al.'s result (explained in detail in prior chapter 3.4.7) where the subjects didn't seem to assess the environment as either safe nor unsafe, was mainly a result of this enhanced response contraction bias, because that study had used a five point answer scale, and that the enquiry should be repeated with different response ranges.

Unwin et al. also repeated a notion from previous studies, that in a before/after study the changes outside the study area, from environmental to social changes, can affect the results between the surveys. Some studies have in fact used control areas to remove such effects.

In their latter report [89], they also reviewed some later studies, and acknowledged that some publishing channels limit or have limited the length of acceptable articles, which may have been a reason why they lacked detailed data necessary for outside reviews with statistical methods, or why they lack descriptions of the actual questions used with the test subjects.

3.7.2 Assessment of studies supporting mesopic knowledge in practical installations, TNO 2010

Triggered by the publication of the Technical report CIE 191:201 on mesopic photometry [3], the Netherlands Organisation for Applied Scientific Research (TNO) released a paper [81] assessing what practical implications it should have on new street lighting installations. The TNO report by Alferdinck et al. includes a comprehensive introduction to the adaptation mechanisms in the eye, to the effects of age on visual performance among other aspects relevant to understanding how the eye works.

In the scope of this study, the interesting part of the TNO paper is the literature review of face recognition studies. They found three studies supporting the idea that the SPD affects face recognition performance, and three other laboratory studies that claim the SPD makes no difference.

Finally, Alferdinck et al. conducted an experiment where 45 observers looked at 12 target persons and gave a percentage rating of how well they could identify the target's face.

There was a stronger connection between the recognition distance and the photopic luminance at the target's face, as compared to the mesopic luminance; this was explained to be so because facial detection is a foveal task and the mesopic model is a good predictor only in the peripheral vision. In the end, the report claims that the effects of lamp SPD were similar between old and young test subjects, which wasn't expected, because older eyes' lenses are more yellow, thus filtering the bluer light giving a higher mesopic luminance. However, this doesn't mention the fovealality of the task. Last, they compared their face identification distance model to the previous studies, and deemed their result data followed similar models, expect for the HPS data from the Raynham & Sansvikrønning study [22] from 2003.

The report authors were able to fit their detection model to the Adrian's visibility model, at a contrast of 0.2. They conclude their report with suggestions for conditions where mesopic photometry and high S/P ratio light sources may be beneficial: lighting objects at the periphery of the visual field, such as borders of the road and maintaining the optic flow of the road layout, and possibly in areas where pedestrians and cyclists are the main users and where they may perceive the bluer light as brighter. With an experience of a brighter environment they may feel safer, although Alferdinck et al. note that there was no sound scientific evidence of such "bluer looks brighter" effect when the report was published.

3.7.3 Number of response categories, Atli & Fotios 2011

Atli and Fotios compared results for one situation when the number of answer options was varied. They had discovered that previous studies had used different response ranges, for example for ratings of SPD, the scales had been from two point yes/no scales to 10 point scales, and sometimes the scale wasn't even reported. [28]

Because literature suggests that the number of answer scale options can affect the results, they conducted a test where 84 students performed simultaneously, on paper, a rating task of four questions of the dimmed lecture hall's lighting and environment. The catch in the study was, that the rating scales were different among the students - they were instructed not to talk during the task, so they didn't know that the scales were different. The scales used had 5 to 8 answer categories. The study was repeated about a month later.

Looking at the results, there was no significant difference in the mean values. However, other interpretations of the data would yield different results when the number of categories was changed. Namely, when the proportion of responses below the middle value of the normalized scale middle point were counted, the proportion varied, for example for the question on brightness it was between 69 % and 87 %.

They then compared the answers on the 7 point scale to the answers on the 6 point scale, by removing the middle category answers altogether. This was suggested in some sources as possibly not having a significant effect on the data. The conclusion was, that with that middle category removed, the distributions of answers on both six category ranges were in fact similar, when statistical tests (Mann-Whitney test and Kolmogorov-Smirnov test) had suggested differences between the 6 and 7 point scales.

With the five point scale, removing the middle values didn't affect the mean significantly, but placed much more emphasis on the tails of the answer ranges, because the middle values had attracted the highest number of answers in most question/response scale pairs. Atli and Fotios take this as a finding in support of the idea that not having a middle value, i.e. forcing the truly neutral respondents to choose either one next to the middle point, they will choose either one at random and the overall results won't be biased. They do call for more studies to validate when and in what kind of research question in lighting the number of response categories matters.

3.7.4 Review of face recognition tests, Fotios et al. 2014

Fotios et al. presented at the CIE 2014 conference a review [109] of eleven studies on facial recognition, and their own experiment measuring the percentage of successfully recognized faces on a computer screen, when the illumination level and the viewing duration were varied.

In their view the worst shortcoming of the previous studies was in the stop-distance method, where subjects walk towards the test target, and stop when they can identify the target. Even at walking speeds the different gait speeds in conjunction with different times needed for the test subjects to make up their mind effectively lead to a random error in the recorded distances. They even refer to a study [110] reporting that changes in lighting can affect gait. Fotios et al. claim that past studies have not tried to define the desirable recognition distance which the lighting should facilitate.

Their conclusions were that facial recognition is a function of viewing time, illuminance, target size, familiarity of the face and SPD.

3.8 Various commercial test installations of LED lamps

There is a major interest among street lighting operators in the operating cost savings promised by the advances in LED technology in outdoor luminaires. Such savings could be realized by power savings, longer replacement and maintenance intervals yielding lower personnel costs. In Finland cities and municipalities operate their own street lighting, and the Road Administration operates the lighting on the state highways. Several parties have already done test installations with LED luminaires, with mixed results before 2010 [90, p. 21–28], [111, p. 59]. Before 2010, the reports from operators mainly reported the achieved luminance evenness criteria and other suitability numbers based on lighting regulations.

Some magazines have published articles on users' and residents' opinions in these areas; the respondents have generally liked the improved colour reproduction compared to previous HPS lamps, but in some cases, the light distribution was considered unsuitable, with majority of the produced light concentrated below the luminaire. In Jaatinen's master's thesis from 2010, she collected experiences from cities' public lighting operators, and reported that some had gotten feedback on excessive glare. The results have been more optimistic in later installations, but these surveys haven't been reported in more detail. As such, their methodology is not relevant for academic research.

The actual questions used are reported seldom in detail, but are given for example for the test site of Myllyoja in Oulu[112], and for a street in Portland, Oregon[113]. In Oulu, the questions posed to users in December 2011 included the following: how much light there was on the road, how much light there was near the road, how was the colour of the light, did you experience glare while walking, how pleasant the luminaire is, and is the luminaire suitable for a residential street. With only 28 responses, their conclusions were that on pedestrian paths the users still preferred the overall lighting produced MH luminaires, because the lighting was more abundant, and the luminaires looked better. On collector roads, i.e. residential roads with some local through traffic, the best rated luminaire was a LED luminaire, surpassing the conventional HPS lamp used as reference[112].

In Portland the questions were adapted from Boyce's parking lot assessment study [18] as Likert scales, agreement or disagreement to the statements for each of the six luminaire types: it would be safe to walk there in daylight and during darkness, in general or alone, the lighting is comfortable, there is too much light and there is not enough light, the light is patchy, light sources glaring, one can't tell the colors of things, the light enables safe vehicular navigation and if they like the color of the light, whether would like that style lighting and how the lighting compares to other streets. The most interesting reported correlation in their study of 38 subjects is that the subjects rated the most uniform lighting as the least patchy; i.e. they were able to correctly assess it visually. This lends some credibility to visual photometric evaluations by untrained members of the public.

Also luminaire manufacturers have sought user feedback, and there are some articles available, but with limited details. For example, the company Valopaa reported in a seminar on end-user centric street lighting in 2012 about some of their experiences. In 2009 they interviewed bypassers, but they only include the results as they had interpreted their data. They found that the colour of light was appreciated as good, but there was noticeable glare. In 2011 they had distributed in autumn and in snowy conditions a survey form to all residences along a relit street; the form could be filled in online. In this instance, the response ratio was very high, but the results are only summarily: distribution and amount of light was "positive" and out of the different CCT's used, the CCT of 4100 was considered the best by the residents. In the same year, they also interviewed users of an automated fuel station, where they had replaced MH lamps with their LED luminaires. Again, we only know that their LEDs were considered better in that they were seen as nonglaring, rendering pure colours and better colour rendering than the old installation.[114]

Based on these finding, we can interpret that those installing new luminaires have sought insight to users' appraisals by interviews and questionnaires. The operators are mostly interested in comparative tests, i.e. whether some luminaires are considered sufficiently better than previous models, so that the possibly higher lifetime costs can be reasoned with the increased user satisfaction, or that operators are not likely to receive many complaints about the quality of the light if operating costs are reduced. Rating scales have included: glare, evenness of illumination, colour of light, colour rendering of objects, and suitability to location.

3.9 Unanswered questions in previous studies

Many of the research papers read for this thesis are concluded with accurate descriptions of possible unexplained or unexplored effects, which should be first confirmed not to exist, before the findings can be validated. We know that we don't know enough to model everything.

The effects behind the uncertainties could be divided into four groups:

The properties of the eye are the autonomous reactions that affect whether the ray of light produces a neural signal strong enough to proceed to the brain. Examples include the size of the adaptation field, the effect of flicker or glare in peripheral vision; there are commonly used equations, but research is ongoing.

The reaction pathways from electromagnetic energy to sensory perception filter and amplify some the signals, with internal and external feedback from the environment and the body. To what extent those feedback systems interact and override each other is not known, we only have correlations for some sets of variables.

The effects of light on affect and cognition have theoretical and empirical backing from lighting, psychology and physiology research, and we know in which direction some variables affect those. There haven't been any reported attempts to describe those as equations tied to any measurable photometric quantities.

The processing of sensory perceptions is not totally reliable, nor always instantaneous. Even if the signal reaches the brain, and an image is formed, the target can go undetected. The effects of attention, sleepiness and related phenomena have their own specialist fields of study, but the light can have a direct feedback to the momentary capacity for processing.

Affective changes beyond lighting are more relevant, for example, for environmental psychology, but the perceived changes in the environment have an effect on the perceptions of lighting - and vice versa.

3.9.1 Effect of luminous parameters and environment on perceived safety

The studies can also be classified on an axis representing the continuum from more complex vision tasks, such as face recognition, to pure vision threshold testing.

The size of visual adaptation field is not accurately known, or how it changes with luminance levels or in dynamic conditions.[91, 92] Most equations of visual performance use the adaptation level as a parameter, but they are based on tests with sufficiently long adaptation periods with, to some extent, uniform visual field luminance. The test setups used so far towards a better understanding of adaptation field size – although involving people as test subjects – have little relevance to real world installation appraisal testing. In the night time traffic environment, the luminances visible to the eye are constantly changing, with values ranging from the scotopic range up to several kcd/m^2 in the luminaires' light emitting surfaces.

A recurring theme in research has been whether the SPD of tested light sources affected the precise phenomenon under study. Older studies had contradictory results, but in latest papers it has been taken as confirmed that the SPD has an effect on vision. For any hypothesised model including vision, the proper measure of SPD and the significance of the effect still needs research. For example, in Knight's research [21] there was no significant difference in the assessments between the two different MH lamps with different CCT, but Ekrias reported [96] that some subjects liked the warmer coloured light better, even if it had poorer colour reproduction. The latter text didn't, however, contemplate whether there is just diversity in people's preferences or whether these diverging responses were due to the interpersonal differences in the impact ambient temperature has on preferred CCT [13, p. 99].

It's undeniable, and agreed upon in the past studies, that in pedestrian environments one of the key tasks is recognizing other people's faces, or their intentions inferred from their faces. Many researchers have used the model from Hall, in which people divide the space around them based on distance into personal space, activity space and vista space, in which we have varying preferences and requirements for comfortable ease of perception. Those preferences follow from the experienced possibilities of interaction, of threat, of escape and from our own reactions.

In the review of several facial recognition tests, Fotios et al. [109] noticed that past research has not tried to define the desirable recognition distance which the lighting should facilitate, but rather has used the theoretically reasoned minimum distance for prompt evasive action as the desirable requirement. This is a crucial question for future studies, because it is not unlike the historic work by de Boer, in which the minimum acceptable luminance levels were reasoned with the possibility of a last second reaction for evading a threatening object on the road. The lighting profession has since come a long way to improve the lighting criteria used, and we shouldn't be content with the pedestrians fleeing at the last possible step.

After that, Yang & Fotios reported of an emotion identification test with six different luminance levels and different distances, with a validated facial image database [115]. At a four meter distance the average rate of correctly detected emotions reached daytime conditions with a luminance of 0.10 cd/m^2 on the test display, roughly corresponding to a illuminance level of 2 lux. At a distance of 15 meters, subjects didn't achieve the daytime recognition rate even at 3.33 cd/m^2 (corresponding to 60 lux at the surface of the screen). An extrapolated luminance level for equal detection rate at 15 meters was 44 cd/m^2 in that study.

The requirements for perceived safety are, naturally, culture and location dependent; in many areas residents feel safe even after dark. For example Kuhn reported [14] that in their study and in both test areas the perceived danger was very low both before and after the retrofitting, which didn't allow for any effects to be attributed to the change in lighting.

Establishing the socioenvironmental cues that signal a perceived need for improved facial recognition (and if the lighting can overcome those cues) could be a welcome addition to our knowledge. We need to establish the relevant cues so that we can know in which locations dynamic lighting control algorithms can be deemed acceptably by the people, and how the dynamic control algorithms' choices are constrained by those cues [44].

3.9.2 Do subjective results measure the intended?

A fundamental question is what attributes can people assess reliably, or if there are aspects of the world where subjective assessments are virtually unrelated to the reality. In usability research, it is acknowledged that only trusting assessments can give wrong results, but it is equally possible to get biased results by only observing and measuring behaviour [65, p. 37]. A combination of methods is required, but the best distribution of effort is situation dependent.

For example, Boyce et al. contemplated [18] that in their study the subjects may have assessed the site for how well they were able to see details, when asked to rate the security induced by the lighting. In another context, when volunteers were asked about a more directly measureable facet of their behaviour, the questionnaire results by Viikari [97] indicated that drivers realize their driving speed changes between lit and unlit roads, just as has been measured to happen by road operators[63].

In a paper from 2001 Veitch [116] explains how subjective appraisals of environments are interconnected with affect theories, but imprecise use of theories about psychological processes is bound to introduce misconceptions and to lead us astray in future research. She continues that there even isn't any agreed model on how to give structure to the interpretations of lighting conditions, i.e. we can't just ask psychology researchers what or which inner measures should we ask people to assess.

In Boyce's often cited parking lot study [18] the answers to perceived safety questions had close correlations between each other, which can suggest that people can't easily assess different aspects of perceived safety (walking there, leaving their car parked there etc.) but rather perceive a single safety assessment. There might be studies to this in environmental psychology, but that – whether people can differentiate between different safety aspects – hasn't been considered in lighting research.

One more unexplored concept is the fact that the subjective responses depend on and relate to different timeframes: some questionnaire items measure the *now*, and others the predicted *future*, or a *longer timeframe*. As an example, glare happens "now", but the uncertainty or fear caused by the veiling luminance relates to the future, i.e. what can become visible when the veiling is removed. Likewise, suitability to subject's own goals is a function of what tasks and experiences the light has afforded, what they can do at the moment, and all steps they plan to take during the near future.

3.9.3 Attributes assessed to date

The subjective attributes used in the studies so far have been numerous. In different countries and languages the precise words and phrases used can't match one to one. Without analysing the possible intents for assessing those attributes, and the facets of the world involved in formulating those attributes, they can not be ordered by relevance.

Looking at all the questions and the number of questions used in the studies reviewed in this thesis, it becomes obvious that there must be some conceptual categories behind them. The same aspects of the environment have been measured with different statements. However, some of them can not be divided unambiguously into distinct categories. For example, the "naturalness of object's colours" is a measure relevant for two categories: in the "direct effects of light" category it is a measure how well the lighting matches the daytime vision in colours, but it could as well be in the category of "world reproduction" as a measure of the typicality afforded by the lighting. The categorization in table 1 is, therefore, somewhat arbitrary - but only to some extent. If one were to consider this categorization when designing a new questionnaire, those ambivalent questions' categories would depend on the motivations behind the study.

It could be noted that the category perceptions of the environment only contains questions related to the prospect-refuge theory and its extensions by Nasar. There are also other proposed generalized measures for environments: Hanyu [117] lists several other variables, also categorized, all of which could arguably be affected by the lighting at the area. Those variables will be reviewed later, in table 2 along with an assessment of whether lighting can effect the assessment of that variable, and the mental timeframe, i.e. whether the variable only depends on what is, or also on what is expected or predicted.

From the studies reviewed in this thesis, the questions asked or items rated can be first divided into 10 somewhat unrelated categories, listed in table 1. These could be further compressed to three groups.

Table 1: Subjectively assessed parameters of lighting in reviewed studies

Category	Measured aspect	Examples of questionnaire items
Overall assessment	Overall grade	good – bad, satisfied – dissatisfied, impression
Light itself	Colour of light	"Like the colour of light": agree – disagree, pleasantness of colour, cold – warm, coloured
	Amount of light	bright – dark, too much – not enough, too strong – inadequate, light, murky, dimmed, luminance on the path
Distribution of light	Luminance beyond the road surface	extensive – limited in area, abundant – inadequate in surroundings & on the verge, concentrated, focused
	Uniformity on the road	even – uneven, uniformity, soft – hard, patchy
Well-being	Glare	glaring – not glaring, just perceptible – intolerable, not dazzling – unbearable, intensive – unnoticeable, brilliant, shaded
	Brightness of the luminaire	
	Comfortability	comfortable, monotony, ambience, mild – sharp, niceness
	Pleasantness	comfortable, monotony, pleasantness, fits the area
Tasks	(Ease of) obstacle detection	
	Face detection	identity of person, eyes & nose & mouth visible – not visible, "quality of the face"
	Face brightness	intensity of the face
	Readibility of signs	
	Visual accessibility	wideness, clarity of place and ease of movement
	Target identification	identity recognition, intent recognition
	Ease of seeing	Close visibility, distant visibility, exposure
World reproduction	Colour reproduction	cannot tell the colour of things, colour reproduction, pleasant colour of skin
	Typicality	natural – unnatural
	Naturalness of vegetation	vegetation seems natural

Table 1: Subjectively assessed parameters of lighting in reviewed studies

Category	Measured aspect	Examples of questionnaire items
	Distance vision	ease of distance assessments, ease of speed assessment, perspective
	Matched to site	Luminaire suitable for this road? Area suitable for recreation?
Perception of environment	Prospect	Good overview? See what is happening?
	Refuge	
	Concealment	Easy for criminals to find hiding places? Few or many hiding places? Chance of criminal hiding here?
	Escape	Chances of escape if assaulted? Easy to get to safety? Easy for a criminal to entrap you?
Inner perceptions	Perceived safety	not at all risky – very risky, safe – dangerous, frightening, worried – don't think about it, feeling uneasy, haste to get away, walking alone
	Alertness	easy to concentrate, make plans, difficult to think, calmness
	Interest, Fascination	Want to explore the area, want to look at the surroundings, fascination, encouraging
	Feeling of extent	Confusing place, chaotic, many distractions
	Compatibility	Being here suits me, I can do things I like
	Being away	Escape from routines, escape from ordinary
Demographics	Traits	femininity, masculinity, anxiety as a trait
Indirect measures	Real safety	
	Energy Efficiency	

The first group corresponds to the behaviour and direct effects of light; how it is distributed, how well the artificial light matches the daytime vision in colours and task performance. The oldest measured subjective appraisals of light (suitable evenness, brightness, glare, colour rendering) were in this group, and there have been at least some studies with a large number of subjects assessing, for example, the glare, brightness, colour reproduction and perceived safety [14, 21, 88].

The second group contains those categories that describe the inner perceptions, from feelings of inner state (e.g. perceived safety) to how the environment feels, for example on the scales of prospect-refuge theory. Some of these items have been measured in several studies, and some dependencies have been identified, but the data is inconclusive. These aspects' assessments are bound to change with time, even in relative short timeframes.

The last group has inner perceptions that can't be changed by the lighting (like masculinity and femininity), or that seem totally subjective. They will not be considered further in this context.

Few aspects studied in reviewed literature were indirect, and best left out of this. We can ask users what they think is the luminaire's light's effect on actual traffic safety of other actors, but the answers are at best considered guesses, even if a slight correlation would be discovered later.

The assessed variables can represent 26 different aspects. These are listed in table 1. Other groupings for the aspects have been used, e.g. in Royer's installation comparison [113], but none of the studies have listed all these variables. In several studies (e.g. [13, 88, 98]) some of these aspects were worth several questions, which were later combined to an index. At least Johansson et al.'s latter article [98] identified that using their previous multiple scale assessment form for outdoor studies is problematic, particularly during the cold season. The resulting Perceived Comfort Quality index from Johansson et al.'s study still has five items to rate, but was shown to have internal and face validity; previous studies had called somewhat similar variables for example "hedonic tone", "pleasantness" or "softness".

Johansson first identified hedonic tone as having the greatest effect on the perceived safety, with perceived brightness being the next most influential parameter, yet in Boyce's data the most significant predictor was the appearance of people, and in the dissertation of Nikunen the colour quality, extent and extensiveness were identified.

Some have tried to give structure to the interpretations, and presented possible models. The results on the effective correlations between lighting qualities, attention restoration theory and common measures for perceptions of safety, reported by Nikunen [13], haven't been validated in other studies. Already the author of that dissertation suggested that areas with crime problems or very low illuminance levels could yield different results.

A similar need to consider the environment and location has been identified in other studies. For example, when Boyce compared a suburban parking lot to a parking lot in an area with cues of safety problems, he reported that the suburban parking lot can attain the same level of perceived safety with a lower illuminance level, if that is deemed sufficient [18]. This means that deriving recommendations from the perceived safety has to take into account the surrounding environment; the luminance level at which better lighting may not be able to help depends on the physical features and the social environment.

There is no agreement on what number of answer options the scales should contain, because it must depend on the question being asked. Even within the same question, there's ample variation in prior practice. For example, in the aforementioned parking lot study [18] some scales were numbered from -5 to +5, others from

Table 2: Variables for environment representation, adapted from Hanyu

Category	Variable	Information Timeframe	Lighting's effect
Particular elements	naturalness	static	enabling
	nuisance elements	variable	can have, or hide
	vehicles	variable	can have
Information afforded and cognitive process	legibility	immediate	enabling
	mystery	future / predicted	can have or hide
	openness	immediate	enabling
	typicality	immediate	can have
	familiarity	immediate	can have
Physical patterns	complexity	immediate	can have or hide, but can also be a source itself
	coherence	immediate	can have
Affective / emotional	pure evaluation	immediate	direct
	excitement	variable	can have
	relaxation	future / predicted	can have
	activity	now & future	can have or hide
	fear	future / predicted	known to have
	interest	future / predicted	can have
Lighting	brightness	immediate	direct
	uniform lighting	variable	has, but is not the sole source[a]

[a] wind, foliage and other moving objects change the uniformity

1 to 7, and some scales had only the adjectives at each end. Others have used 4, 5, 6 or 9 point scales from the value 1 upwards (labeled with words at each end, for example Atkins et al. [79]), but also a scale from -2 to +2 has been used. Some have tested on a continuous scale, i.e. a line where subjects would tick a mark between the endpoints (e.g. [94]).

In a methodology review at least seven other prior studies were identified by Fotios, studies in which the subjects had an inconvenient tendency to mostly use the middle values of a seven point scale, to such extent that the differences in appraisals were smaller than expected and no conclusions could be drawn for those questions [30]. None of those studies were reviewed in this thesis.

3.10 Summary of findings

Before 2010, many lighting studies involving users and their subjective impressions often lacked critical statistical evaluation tests to the significance of their findings [15]. After that, the majority of reviewed research (e.g. [56, 89, 100]) has included at least some figures beyond the difference of mean values.

Luckily, though, multiple researchers have shown similar results, so that we do not need to view their findings as purely indicative of the direction of the effect. However, only with more studies and more comprehensive data analyses we could conceive a trustworthy predictive model of these effects.

Lighting can accentuate both positive and negative feelings about the environment, but it can also influence where people focus their attention, as well as affecting how people behave.

Factors known to have positive effects on perceptions of dark time environments are:

1. visible greenery [13]
2. non-glare
3. evenness

Factors of lighting and factors that lighting can alter, and which are claimed to affect safety appraisals are:

1. brightness [18, 88, 117]
2. perceived appearance of people, both of the face and of the whole person
3. naturalness, pleasantness and monotony of the light [88]
4. entrapment [77]
5. concealment [44, 77]

Most reported studies have acknowledged, that the test setups can affect the results adversely, so more studies are needed. At the very least, the laboratory results need to be verified in real environments. That is in part because in laboratory experiments the test subjects are usually, if not always, not subjected to other vital sensory inputs, like sounds [13, p. 89]. Excluding other senses is likely to affect the performance and cognition, but the direction of the effect can not yet be predicted.

It's evident from the reviewed papers, that lighting studies typically take a long time to complete. Several studies were conducted over a course of several years (e.g. [17, 18, 21, 79]).

Several studies indicate that brighter and more evenly lit environments are considered safer and preferable, for example those described by Boyce [18], Johansson [88] and Hanyu [117]. Based on other studies (e.g. Rantakallio [20] and Nikunen [13]) it's arguably plausible that the connection of brightness and perception of safety diminishes once some luminance level is reached. In Johansson's study of individual and personal factors affecting pedestrians' perceptions, the assessment of lighting could explain 45 % of the variance of perceived danger in their data. This would indicate that the quality of lighting is a major factor in providing a feeling of safety, even if half of the assessment is made purely based on the features unrelated to the lighting.

However, that is not the final answer to improving lighting, as other studies have found out that excessive brightness might not even be viewed as pleasant,

Table 3: Effects of variables on perceived safety from previous studies

Variable	$\lvert Beta \rvert$ [a]	N	Study
Hedonic tone[b]	0.38	81	Johansson et al. 2011[88]
Entrapment	0.37	122	Blöbaum et al. [77]
Prospect	[c]	50	Haans and de Kort [56]

[a] absolute value of standardized beta-coefficient; direction of any scale varies between studies

[b] index, variables include: pleasantness, brightness, monotony

[c] analysis used unstandardized units, prospect was by far highest

nor relaxing [117]. Brightness, when perceived as excessive, along with glare, can reduce two components of the ART: fascination and being away. The result was not yet statistically significant, but had support from the free responses in Nikunen's study [13, p. 92].

Given the high number of subjects in some studies, it's a shame that they lack some of the statistical analysis methods that could have been used. For example, the study by Kostic and Djokic had 112 subjects, found significant differences between the light sources, which answered their research question, but the report didn't provide any regression equations. They almost had enough volunteers to fulfill the rule of thumb cited by Veitch [116] that tests should have a minimum of ten subjects for each scale assessed by them for a factor analysis, so even if the condition was unmet, values could have been useful in selecting new research questions. Without factor analysis, their study doesn't tell us which of the other questions had the greatest effect on, for example, the perceived safety.

Despite the results of Haans & de Kort's study [56], in which walking pedestrians preferred that the light budget was allocated in such a way that their immediate vicinity always had the brightest luminaire (reported as mostly affecting the concept of prospect), it is plausible that the effect was a result of different *perceived* overall lighting level. This claim follows from the fact that the contribution of the more distant luminaires to the overall perceived brightness is nearly negligible, even so that the luminaires outside the test area contribute to the perception more than the tested luminaires. The distant test luminaires could, despite the reported results, in fact contribute to the response cue, but the test could be repeated with a constant *perceived* lighting level for validation. With unstandardized units, the measurable effect in their report can not be compared to the beta-coefficients reported by Blöbaum et al. [77] and Johansson et al. [88] (Listed in table 3).

More user studies have been conducted with indoor lighting, but their results may only be very carefully applied in the outdoor realm. This is due to the fact that not only the context is so different, but also the brightness contrast, viewing distance, private-public context, time of day, effect from other nearby luminaires, and the illuminance levels are vastly different, so even with same subjects, the users are not the same. We can compare this to usability engineering, where the two major prerequisites to improved task performance and user satisfaction are knowing the user and knowing the task [118, p. 22]. The visual and social tasks vary, and the light has an effect on the road user's cognition and visual performance, which can only mean that the predicted appraisals depend on different variables, and with different coefficients.

What can be perceived as natural, may well be learned expectations, either from the modern era with a long history of incandescent lamps, or even deeper in the evolutionary roots of mankind, where low colour temperature light generated with fire meant safety at night [13, p. 94–95]. Recent studies have indicated that at the indoor office luminance levels a bluer light of higher CCT can enhance the alertness in humans [12, p. 63], but no studies of that effect in mesopic conditions was found. Higher CCT light has also been shown to raise cortisol levels in high school students in the dark winter season [119].

In mesopic vision, the human eye is more sensitive to the bluer end of the visible light spectrum. In practice, this has been seen, and even proved, as a indication that luminaires with higher correlated colour temperatures yield the same perceived brightness and visual performance with smaller radiation output. The practical measure of this effect in a light source is the S/P-ratio (short for scotopic/photopic ratio).

However, there is little support in the reviewed studies to the provision given in some countries' road lighting standards to lower the illuminance levels when light sources of high colour rendering index (R_a) are used. Results do support that such power saving measures are justifiable, but the criterion is some other measure; possibly the S/P ratio.

Moving the emitted power more to the shorter wavelengths is not without challenges. There does not seem to be any conclusive research to the adverse effects of bluer light on the natural environment and animals, compared to the benefits of bluer light. Bluer light can cause disruptions to bird migration navigation and animal reproduction practices, such as the hatching sea tortoises [34, p. 221]. Some bluer LED sources have a high power emission peak in the spectral band of 440–460 nm, for which the human retina has a lower threshold of light damage than it does for longer wavelengths [120].

In other words, we do not know, if such light with relatively more power in the bluer end of the spectrum, more daylight-like lighting can be utilized at so much lower levels that the stronger negative effects are counterbalanced, or whether it would be more practical to use specially designed light sources, e.g. with lower CCT, in identified sensitive natural areas. Exposure to light at night is also known to affect human health already at low levels: it decreases pineal melatonin production and secretion, which affects the circadian rhythm. The exact relationships between the SPD, illuminance level and this adverse health effect is still under investigation [121, p. 2715].

There is some evidence that people seem to gaze at the highlighted objects in their vision more often than at the subdued objects. In Cengiz's report drivers mostly looked at the area lit by the car headlamps at night, but gazes were spread out more in the daylight condition [91], and Caberletti reported that highlighting the centre console helps drivers find the controls positioned there [94] . We are attracted towards light, but too much uneven light at one spot causes dissatisfaction [117].

3.11 Conclusions from prior research

Out of the lighting studies reviewed, several were even described as exploratory research. This would indicate that holistic study of user appraisals of lighting seems to be in a preliminary stage, where the problem isn't clearly defined. It has been agreed on that the subjective assessments should be improved, and many subjective attributes have improved in the studies, but their internal order isn't known. This is not a problem for applied research, i.e. partial information has been sufficient for improving the real world use cases.

Many parts of the holistic problem are, however, clearly defined, but mostly they have not been combined into hypotheses, which could be rejected or confirmed. Given the fact that in these studies so many variables have been identified as affecting the appraisals (including properties of the physical and social environment, personal traits and photometric quantities), and the fact that only confirmatory research should be used for generalizations to the population at large, it can be said that it is about time a validated multidisciplinary qualitative research methodology for affective and cognitive effects of lighting could be developed and formalized.

Faced with these aspects and all the questions used so far, we get back to the first research question of this thesis.

There is a strong connection between brightness and perceived safety, but the luminance of the scene can be measured without the users. Even if the perceived brightness does not depend linearly on the luminance, the mapping from luminance to perceived brightness is monotonic.

Users have been asked to state and to order the important factors for lighting, and in the various studies the second most important factor has been occupied by evenness, pleasantness, quality of light and lack of glare (Table 4). Out of these, other studies have shown factor loadings in support of pleasantness, softness (being closely related to, or the same as evenness). Other items found to have an effect are calmness (also titled enabling relaxation) and fit to area.

When we compare the items mentioned in the previous paragraph to the theoretical foundations of environment assessment by Hanyu, and those referred to in Nikunen's work establishing a link between ART and lighting, we find that they have one item in common, and for that item to be assessable in the dark time, the lighting is needed. With Nikunen's result of natural elemens correlating with assessments of a relaxing environment, and combining that with the items naturalness and relaxation in the table 2, we can propose that the naturalness and relaxation measure the same property, but in different timeframes. Naturalness was also an important factor in Johansson et al.'s perceived comfort quality index (PCQ) after (plausibly) technically measurable parameters of the lighting.

Before we claim the choice final, variables identified in other known studies should be considered. In Rantakallio et al.'s report [93] the most important items for perceived appreciation were fit to area and that the light makes the environment look pleasant, followed by enabling relaxation (see table 5). Pleasantness and calmness were listed by Viluinas et al. as the two most important variables in the same context, but Johansson's previous work equates pleasantness with the term hedonic

Table 4: Subjects' self reported most important factors for lighting, from previous studies

Report	overall or choice[a]	Most important factor	2nd important factor
[90]	overall	amount of light	equal: lack of blackspots, no glare
[93]	overall	quality of light	amount of light
[21]	overall	brightness	
[20]	overall	brightness	pleasantness
[100]	overall	feeling of safety	feeling of comfort
[100]	choice	feeling of comfort	illuminance uniformity
		Aspects mentioned often:	
[20]	choice	evenness, brightness, lack of dark spots, fit to area	

[a] here "choice" means selection between presented luminaires

Table 5: Effects of variables on perceived well-being from previous studies, sorted by effect

Variable	factor loading	N	Study
Fear of assault	0.89	11	Viliunas et al. [78]
Pleasantness	0.87	21	Viliunas et al. [78]
Sharpness	0.856		Johansson et al. [98]
Calmness	0.84	21	Viliunas et al. [78]
Small object visibility[a]	0.79	11	Viliunas et al. [78]
Encouragement	0.77	21	Viliunas et al. [78]
General perceived safety	0.76	11	Viliunas et al. [78]
Niceness	0.75	21	Viliunas et al. [78]
Fit to area	r=0.82	55	Rantakallio et al. [20]
Light colour makes the environment look pleasant	r=0.70	55	Rantakallio et al. [20]
Relaxing and being away from stress feels easy	r=0.64	55	Rantakallio et al. [20]
Skin looks pleasant	r=0.56	55	Rantakallio et al. [20]
Easy to recognize intents of people	r=0.56	55	Rantakallio et al. [20]
Good perceived safety	r=0.56	55	Rantakallio et al. [20]
No glare	r=0.27	38	Royer et al. [113]

[a] in the paper classified as a cue of prospect

tone and with the PCQ.

The aim was to identify the single most important subjective item, or two items, for perceived danger. Having compared the concepts used in previous studies and their analyses, the subjective item *naturalness* has the greatest effect to the perceptions of safety, after brightness and lack of dark spots, both of which can be measured without users. All the studies on which this assessment is based on, were conducted in relatively safe urban and suburban areas. In areas considered less safe already in the daytime, a different item could be more important.

One has to ask, though, what does "naturalness" mean? The word in itself has many meanings from abundance of greenery to being an object typical of its type, and the question to pose must be unambiguous. Hur et al. explored the connection between appreciation of areas and several variables including naturalness, and reported that the perceived naturalness is not directly a function of amount of tree cover, but influenced also by the openness of the environment [122].

This openness gets us back to Appletons prospect-refuge theory, where an open vista relates to good prospect, and preference. Studies and correlation coefficients listed in table 6 indicate that the natural reproduction of people and colours, and the extent of the lit area were previously identified as the best predictors of perceived safety. Combining these notions, it is safe to say, that the naturalness refers to the extent to which the light reproduces the environment and persons in it like in the normal conditions (i.e. it looks natural) and the extent to which the light allows people to have a extensive, i.e. open view of the area, like they do in the daytime.

The Likert scale statement proposed in this thesis is then

- "The lighting resembles natural daytime conditions" for pedestrian environments.

It does not directly predict the perceived safety, but is a major coefficient in the equation. This even binds the perceived safety to the aspects typically considered intimidating in daytime environments: unnatural elements and blocked prospect. The statement looks trivial, but is backed by the findings of previous studies reviewed in this thesis.

Table 6: Correlations between perceived safety and other items from previous studies

Item	correlation coefficient[a]	N	Study
Brightness	+	356	Knight 2010 [21]
Brightness	+	55	Rantakallio et al. [20]
Brightness	+	15	Boyce 2000 [18]
Appearance of people	0.77	15	Boyce 2000 [18]
Colour quality	0.58; 0.72	29; 26	Nikunen 2013, study 4 [13]
Extent	0.64	26	Nikunen 2013, study 3 [13]
Extensiveness	0.60	29	Nikunen 2013, study 4 [13]
Being away	0.52; 0.60	29	Nikunen 2013, study 3 [13]

[a] + or - when only the direction is reported

If assessing two scales simultaneously is later deemed possible and convenient, the proposed statements for a Likert scale are

– "The environment appears natural under this lighting" and

– "The lighting enables good perception of the area."

3.11.1 Questions derived from empirical evidence

From past studies we can infer some possible directions for future studies. These include findings supported by the evidence, but which have not been validated with for example hypothesis testing. Only then can we compare and order the identified effects with regard to their importance, with all due certainty.

Excessive brightness If further studies can confirm the findings of Nikunen [13, p. 92], that there can be too much brightness in outdoor environments, this could advance the acceptance of light pollution reduction measures.

Highlighting greenery Given the finding that light distributions which spread out the light to greenery elements were favourable [13, p. 74], and because the theory behind how that added light on those elements has positive affective results, luminaire designs easily enabling such installations could have positive effects on the acceptability of lighting changes.

Several studies indicated that lighting the areas outside the road are beneficial for visual performance, and for the appreciation of the lighting. Whether this increases the appraisals of naturalness could be verified with users.

Acceptability predicted by entrapment and concealment Boomsma & Steg reviewed several studies on the effects of concealment, entrapment and other attributes of the environment on the perceived safety. Based on their review, they proposed that entrapment and concealment have the greatest effect on the minimum lighting levels perceived as safe and on the acceptability of those levels.

One must remember that in dark times, lighting can minimize entrapment in some environments: adding light makes more escape options visible. For visual performance, Rea's visual performance simulation showed that when the ambient illuminance is very low, lighting only the immediate vicinity of an intersection provides little benefit except for targets near the intersection itself [45]. This could be supported by the relation between lack of concealment and perceived safety and preference. If the properly lit area encompasses only the task area (the intersection), which only forms a small part of our vision, at least people with a trait of small environmental trust get a feeling of uncertainty because of the lack of information of what is happening inside their action space, inside which the task area is situated. In other words, a road user approaching a lit intersection in an otherwise dark environment does not know what hazards are present in the darkness nearer to them than the lit intersection.

Directing gazes with highlighting Some of the eyetracking studies have indicated that people seem to gaze more often at highlighted objects, than they do at objects of lower luminance in their field of view. It would be interesting to be able to analyze the recorded gaze directions from all known studies in this regard. Even if the effect seems trivial, nobody has proposed a function for predicting the allocation or probabilities of gaze directions as a function of the various relative luminances in the scene.

3.11.2 Test designs

It is evident from the reviewed literature that several past studies have neglected some of the possible sources for errors, as listed in chapter 1.1. When planning future studies, the following points should be considered in test design:

Range bias is arguably present in many prior studies. It is easily introduced, when testing physical stimuli on a preset range. Methods that can be used to mitigate it's effects are to use different but overlapping ranges for different subsets of test subjects; then there should be a difference in the variance between those subsets, if one of the subsets' range included a significant threshold. A simpler method is to test a larger range of stimuli, and to employ more test subjects.

In many cases, presenting the extremes of the stimuli range before testing can be used to anchor the subjects' perception range prior to any assessments. The validity of this method has to be evaluated beforehand, because it can introduce priming effects.

Priming effects should be avoided. These are when the subjects are, prior to their test, exposed to words which then stimulate related concepts in their mind, which affects their thinking. An example used in lighting research has been that semantic priming occurs when the instructor asks test subjects to rate the effect of brightness on their personal safety, the subjects will to some extent agree that lighting has an effect on their personal safety, even if they had not thought about the connection beforehand. Had they been asked to simply write down factors affecting their personal safety, their thoughts would not have been altered by the question.

Avoiding semantic priming totally is not always possible when examining concepts that require an introduction of the concept to the volunteers. Examples of such could be assessments of the components of ART, although they have usually been measured indirectly: subjects have assessed more concrete features of the environment, and those appraisals have been combined to indices corresponding to the components of ART.

However, when examining the resulting index or indices, it is important to acknowledge that when a model is built on one data set, and the predictions then compared to the original data, there is an inherent selfcorrelation. This means, that to prove that each index really measures the intended complex variable, i.e. the validity of the each index, the effect of selfcorrelation has to be calculated and removed, if the validity is tested on the full answer dataset that is used for the study question.

Another way to mitigate the priming effects is to ask more scales than necessary simultaneously, so that the questionnaire stimulates several concept pairs. The effect of priming is then randomly spread evenly across all subjects and combinations of concepts, given enough answers.

Counterbalanced order for the tests is needed to remove any learning bias or other effects. When test subjects are sent on a route, some of them should walk the route in the other direction, when possible, and if a comparison run is conducted in the daylight hours, the order should be reversed for half of the subjects. In a more elaborate case, when the test includes samples at varying distances, the tested distances could be seemingly random, i.e. not sequential from farthest to closest (or vice versa). In some test setups it might however be impossible to counterbalance every variable; the choice of parameters not counterbalanced should then be reasoned for the reader.

Control group If one only compares a change in appraisal after a change, or differences in appraisals between two conditions, for example the time difference and the learning effect can swamp the real difference in the users' cognition. Therefore it's imperative to use a control group for validation. In simple cases with a short time span, though, the counterbalanced order can already provide the control group within the subjects. A control group's data can be also used to identify invalid recordings; sometimes one subject's data might be technically incorrect, and should be removed from later analysis.

Focus group interviews, as in the subjective appraisal study by Kostic & Djokic[100], help make the questions better understandable for the test subjects and reduce ambiguity.

Uniform conditions when several locations are used. This includes not only physical curvature of the road, distances between luminaires, installation height and other physical measures, but also perceptible differences, i.e. differences in how the subjects perceive the areas. For example is the place high in concealment, or is everything in open view?

The studies listed in chapter 3 have used several attributes' assessments (beyond photometric measures) to describe the differences between the environments. From those listed in table 1, the assessable aspects of entrapment, prospect, extensiveness beyond the road area, comfortability and pleasantness, typicality, match to site and prior perceived danger have been shown to have some effect on the other assessments, so at least some of them should be included when checking for uniform conditions.

Prior perceived danger has been so low in several studies (e.g. [13, 14, 56, 90]) that there was no noticeable difference in perceived safety between the before and after questionnaires. This is especially true where the concept of safety is presented as a lack of physical, verbal or sexual assaults, but where the occurence of such crimes is virtually unknown to most residents. In my view, this point would need some conceptual work together with the environmental psychology professionals' community, to further define and narrow down the concepts involved. Only then we could know what are we asking from the test subjects, and what should we be asking. There are necessarily differences between areas, for an extreme example between crime ridden neighborhoods and rural areas with a frequent wolf or bear sightings.

Sufficiently large differences in luminance or illuminance values should be used, if it is expected to have an effect, and is not just a byproduct of the compared installations or luminaires. For example Boomsma & Steg [44] only compared 12 lux and 17 lux conditions; although they did find some effects, the authors themselves reported discontent with the number of values and the range. This is different from the previous point of prior perceived danger in the sense that these can be controlled by the researcher, whereas the perceived danger can not be controlled beforehand, but can only be assessed during the tests.

Number of subjects has often been too low. It is widely known that recruiting test subjects can be hard, and running the tests requires more effort when the number of subjects increases. Calculating the confidence interval, statistical power or other indicators for hypothesis rejection requires knowledge on the sample and on the options, and the nature of the applicable distribution, so presenting an equation or valid numbers is beyond the scope of this thesis. The fast rule cited by Veitch is a good start, anyway: for factor analysis, for every scale assessed by the subjects, the test should employ a minimum of ten subjects [116]. If the significance tests fail, it is good to have a backup plan for running the tests with more subjects.

It has to be acknowledged that recruiting more subjects involves a higher workload and paying out higher compensations for the subjects. Nevertheless, the work expended during an actual test situation is a fraction of the total effort in designing, conducting and reporting a good study.

A representative sample of population is a prerequisite for generalizable results. The representativeness has been addressed in some studies (e.g. [14, 44, 77, 97], at least for some parameters such as age). Other studies (for example [8, 16, 18, 20, 21, 56, 93, 98, 99, 100] have reported the limitations of the sample (age range, gender) without directly mentioning that the representativeness is not statistically analyzed – this is however usually a direct consequence of the small sample sizes, or a trivial result when the subjects represent only a tiny age range, such as volunteering students.

A direct effect of age on vision is the yellowing of the eye lens in older people, which reduces the amount of light reaching the retina, and alters the spectral distribution. The personal traits, such as masculinity and feminity were shown in Blobaum's report to affect the appraisals of the safety and others [77]. Also professional knowledge of lighting, of traffic or safety or of built environment influences the appraisals. These variables should be considered and tested for, if generalizing the results to the whole population.

In Knight's study[21] there were also minor identifiable differences between residents of different countries, i.e. only in some countries a statistically significant correlation was evident for some variables, but in general there are no known major differences in road lighting preferences between residents of different countries.

Number of response categories for optimal results is dependent on the research question and the attribute being tested. Literature [28] suggests that when it is more important to isolate those respondents who do not know which way they would choose from the middle point, from those who have an actual preference to either direction, one should use a scale with an odd number of options. When it's more important to get an indication of the direction into which the study population is leaning towards, omitting the middle option would give a more representative distribution.

If the number of different stimuli is higher than the number of response categories, some different stimuli unavoidably get identical responses. If two nearly same stimuli are viewed with a long time interval, their order in the assessments may be unintentionally reversed. In very large response sets the average answer should be different, but this counterbalancing method is often unfeasible.

This is also relevant when the test subjects compare consecutive installations, they must decide whether their perception of difference from the previous installation is large enough to warrant answering with the next value on the scale. With more than 7 ± 2 effective chunks of information to remember, the past assessments are no longer in the working memory, and their expressiveness is reduced. It was beyond the scope of this thesis to review whether there has been any research into the encoding capacity of brain when the task is to specifically remember the qualities of lighting. A laboratory test setup could be used to analyse how many consecutive scenes' attributes people could remember reliably, with different number of attributes assessed for each scene – the number of test subjects and repetitions would have to be high, though.

A practical way to address this challenge would be to always use a pilot appraisal test, where only few persons' appraisals would be compared with expert appraisals. If the answers don't show an expected difference, either the question or the number of answer categories is amended, in hope that the next pilot group finds the difference. It's also possible subjects simply can't perceive the expected difference.

Highly variable CCT between different lamps, when visible simultaneously, has affected the appraisals in previous studies [36]. Unless the aim to the study is to compare available luminaires, or to explore the differences attributable to the CCT, or the perceptions of different CCT light sources, the CCTs should be kept similar.

Multidisciplinary team of researchers. Some of the simplest fundamental questions of vision, detection and light may have been answered to an adequate degree, yet new simple questions still surface with changes in viable light sources and the built environment. However, when the performance, cognition, biological effects, affect and preferences of people are studied in relation to the lighting, no single discipline can know it all; even less so any practitioner of a single discipline.

Incorporating the needs and interests from other disciplines is likely to yield more complex tests and larger sample sizes, which means increased costs and longer realization times for a single study. It is no longer a question whether it is worth the effort, but an identified need for advances in many fields. For example, geographical information research has noticed the increasing need for crossdiscipline knowledge and co-operation, for modelling the location dependent data in effective ways [123, p. 280].

The Living Lab approach is a possible candidate for addressing several of the factors listed above. In a LL the stack of research data is constantly added to, and the data points are tied to the locations. Then the environments can be identified, which have the largest amount of data points, or which are significantly different from other locations on any indicator, or those most considered an outlier. Those environments can be later assessed when a new research question is formulated. In a successful LL the pool of subjects would be available to crowdsource any new required subjective assessments.

3.11.3 Requirements for reporting

When reports of user surveys and user experiments have been later analysed together with other similar studies, research has shown that the reported data is surprisingly often incomplete. Consequently, too often the reported findings can not be generalized or validated conclusively. Such results have some practical value for lighting designers operating in similar areas, but researchers can't blindly take them as prior knowledge.

Old studies seldom report conclusive data required for assessing the statistical significance of the findings. For the oldest studies reviewed here, it should not be considered a failure by their authors, given the burden of computer history. It is however noteworthy that many modern research papers could still give more detailed figures; this could help others when they seek to compare the results of different studies. Despite that suggestion one must remember that the members of the research community should be able to trust others to conduct their studies with an appropriate level of rigour and attention to procedure - the meticulously detailed statistical tests are still secondary to the results pertaining to the effects of lighting.

Excessive data in reports might not serve any function. For an example, Iwata [84] reported a high correlation ($R^2 = 0.9071$) with the preocular illuminance (at various distances) and the visibility assessment, yet the dependence follows from their test setup with only one luminaire and the subject moving towards the light source. The note in the text can only lead the reader astray into thinking there could be a causality effect between preocular illuminance and visibility assessment, when the distance was the cause for the increasing preocular luminance. Had the focus been more on a factor analysis, such that they would have been able to indicate what explains the difference of the correlation R^2 from the value of 1, the figure would have been more relevant for the readers.

Research papers on street lighting user experiments should report:

The properties of the before/after installation Technology changes sometimes fast, even if the typical lifetime of a road lighting luminaire is 10 to 30 years [52]. What seems obvious at the moment, probably isn't anymore if the results need validation after 10 years – and it's concievable that it may take a decade from the first findings to a validated model.

Even if the luminance levels as such don't matter in the planned research, later researchers might find your data usable for novel data analyses later. Things to consider are for example measurements of the luminances and SPD, drawings of the site and measurement points, photos with imaging luminance meter.

Obviously, not all journals will accept details excessive to the publication's results. Even then, the measurements could be publicly stored online. The data from previous decades works probably won't become available (because of the needed work effort, i.e. money), but for future publications, this could be beneficial.

Even if unexplained results could not be explained by later analysis, even when accompanied by extensive data on the photometric measurements, such datasets could be revealing for posing new research questions and for formulating new test setups.

The demographics for the subjects. Sometimes special considerations have to be made for those with an impairment; poor eyesight, mobility problems.

Any weather effects For example, whether there was snow cover, or if there was rain on some nights. Rain diminishes the contrast of the scene, and snow increases the luminance to levels only attainable with significantly higher lumen output. Global readers can not know how to describe the prevailing normal weather in every country.

The exact questions used are paramount for later comparisons.

The numerical data usable for later statistical analyses. If possible, the raw data could be available online, outside of the publication.

The important findings in a format suitable for the data; when possible, this means equations and graphs. It's worthwile to acknowledge that data visualisation for complex interactions is an art of its own, and as such beyond the scope of this thesis.

The graphs and other visualisations that help assess and interpret the results. If the data is not relevant for the results, the presentation should be considered carefully and at least the important graphs are best presented positioned away from the unrelated visualisations.

4 Product development of a feedback device

In line with the goals set forth in the introduction, this section will describe the device designed and constructed for an easier and more effective feedback collection process during outdoor lighting user tests, and more importantly, the process used, and the reasons for the choices made during the development.

4.1 Goals

After and during the literature review, the goal was refined according to the findings. The feedback system should, to any reasonable extent, be usable for two similar use cases in different context:

1. Collect user appraisals of the given subjective attribute in real time on a test route
2. Collect user's indications of good or bad locations relative to a given subjective attribute in their daily lives, i.e. in a Living Lab environment

If the contexts set conflicting requirements, the first context should be considered as having priority, because such test routes are faster and easier to set up and therefore the number of uses would be higher, until a suitable LL has been set up. In both contexts, the goals for the recorded data and for the user experience were initially set to be:

1. All stored data points shall be tied to a location for later analysis.
2. The user must not need to wait for the device to operate it; waiting would require the users to shift their concentration from the environment to the device, which would likely affect their perception of the environment.
3. The battery must last at least long enough for one study night with several participants on a set route.
4. The device must have an identifier both printed on it and available wirelessly, in case several devices are operated in close proximity.
5. The device should have an option to remind the subjects, by vibration or sound, if they seem to have forgotten the task to give their appraisals at regular intervals, even if it hasn't changed from the previous.
6. The user interface should be such, that the device could also be used in vehicles and while driving vehicles.

Initially, the data collection was envisioned as recording only quantitative data. In a focus group discussion with researchers possibly using the device in the future, a recurring concern was that the qualitative data of the test subject's experience is essentially required a proper analysis of the lighting in question. The statistical methods that can be used to analyze the recorded appraisals, don't prove causality, even if some methods reveal mediation effects. The quantitative data by itself can be hard to interpret in the analysis phase, especially so when the unconstrained parameters affecting the study are not known. Hence, one more goal for the data collection process was added:

7. To facilitate collection of qualitative data, the system should also allow textual questions to be presented at times, if later deemed necessary, or it should allow the study instructors to ask those questions immediately after the set test route is finished. To link the questions to the appropriate location, the instructor should be able to visually assess the recorded data immediately at the end of the test route.

4.2 Requirements

For the device to be usable in future lighting research, the output should be validated in the sense that the data produced is correct, is of value to resarch, and accurately represents the test subjects' sentiment of the survey question. Also, because the vision sense is adaptive and the vision performance is context dependent, the device must not influence the adaptation level, nor the factors involved in any identified models of users' mood - to the extent practical for the chosen technology. With these top level goals in mind, the technical requirements can be classified into two groups: those affecting the user experience, and those defining the minimum technical perfomance level.

The goals and requirements above were used to define a list of technical requirements, implementation details:

1. The input method must be usable with just the new device and a mobile phone or a tablet computer.

2. Test subjects must be able to give their feedback with gloved hands, if and when it is cold outside.

3. The phone does not need to have data connection when the survey is conducted, but if it has, the logged data can be sent immediately.

4. The battery must last for four hours, but preferably longer.

The trivial solution, current day smartphones with touch screens, as the sole input method can be excluded on the basis of their screen brightness, and because they usually require users to handle them ungloved for the touch screen to register the interaction, which is a problem in cold climates. While it may seem trivial to remove gloves to operate the touch screen, the planned recording sessions are supposed to last long enough that in the cold weather it would cause at least discomfort, if not frostbites, and because of that, the test subject's mood and attention target would be affected uncontrollably.

The prevalent luminances of surfaces found in outdoor night scenes are in the mesopic range, from 0.01 cd/m^2 to 10 cd/m^2, and occasionally even up to 40 cd/m^2 (ignoring the glare sources). For most smartphones the various sources identify the minimum luminance (of a totally white screen surface) at 50 to 80 cd/m^2 range [124], which is way too high in lower mesopic surroundings, especially because the graphical user interface would then have to be brought near the users eyes, thus covering a considerable portion of their field of view, possibly even causing glare.

With such large differences in luminance, one also has to consider whether the subject's adaptation level could be affected. There are no universally valid equations for predicting the adaptation level, or how nonuniform luminances in the field of view affect it. However, a reasonable estimate that has been used is that an angle of 20° around the fixation point should be considered - at least if it's of even luminance. The human iris can contract fully in 0.3 seconds and dilate back to fully open in about 1.5 seconds, providing a 16 fold decrease in the amount of radiation reaching the retina if the autonomous nervous system finds it necessary - total contraction won't, however, happen until daylight luminance levels are reached. In addition, synaptic adaptation occurs in about 0.2 seconds and can account for 2 to 3 log unit changes in the scene luminance. Larger adaptation changes require slower adaptation mechanisms in the retina. [25, p. 63]

Based on the previous paragraph it is argued here that in the worst cases, where the prevailing luminance level was 0.01 cd/m^2 and the mobile handset screen had not been set to lowest possible brightness levels, a screen surface with luminance of 100 cd/m^2 could occupy a significant area within the adaptation angle referred to above. If the screen occupied the 20° angle mentioned above fully, i.e. a space angle of 400 square degrees or 0.0049 steradians (for example, a 4 inch mobile phone screen at a distance of about 28 cm, confirmed as a usual viewing distance with two subjects), and the screen was turned on suddenly, the eye could adapt to 10 cd/m^2 within the first 0.2 seconds, and at the time point 0.3 seconds the adaptation level could reach 160 cd/m^2. This is sufficiently higher than the screen luminance, that even if the iris did not contract fully, the actual adaptation level could be close to the screen's luminance, assuming the 20° angle assumption holds.

After the appraisal was given, if the screen were to turn off, the adaptation level would reach the original state in the 1.5 seconds, and almost the same level already in 0.2 seconds. In reality, though, the lit screen would be moving with the hand from the periphery towards the fixation point - and the fixation point towards the hand - so that the upward adaptation could be nearly complete by the time the screen is in it's final position. It is more likely that the movement of the hand away from the direction of view would not affect the readapation to the scene, because regardless of whether the screen turns off instantly or midway to the peripheral vision, it disappears faster than the iris can react, but at a rate comparable to the synaptic adaptation timeframe of 0.2 seconds. At a greater viewing distance, the adaptation level would be affected less, which would, however, increase the chances of discomfort glare.

In the end, even if the use of a handheld touchscreen would only affect adaptation levels for a short period, it could nevertheless be a source of error in the appraisals.

At the very least the screen brightness settings should then be set to similar, constant levels on each test run and on each used device.

For verification purposes, the luminances of a typical low end smartphone and a tablet computer screens were measured with different screen contents and brightness settings, in a darkened room with the Konica Minolta LS-110 spot luminance meter; results are in table 7. Even if the brightness settings could be reliably set before and for the whole duration of the study, in practice the user interface would unavoidably cause sensations of glare at least in some situations. The effective colour resolution of such small screens is not known to scientific precision, but is reasonably reported as "effectively 64 thousand colours", or 16 million colours for some devices. These figures correspond to a colour space of 16 bits and 24 bits, which leaves 5 bits or 8 bits for each colour.

In the stimulus range from 0.1 to 10 cd/m^2 there are about 200 steps of just noticeable difference, and the curve follows a logarithmic shape [125, p. 424]. Such smallest noticeable differences do not guarantee an easy, fluent and speedy identification of the screen contents. When we also know that the pixel value to luminance functions are different for the screens of various devices, to guarantee an accurate perception of the user interface, only 32 levels of grey could be used. Given that for the worse performing device out of the tested, when forced to a minimum brightness setting, an all-black screen has a luminance of 0.09 cd/m^2, the first usable grey level that could be utilized would have a luminance of 0.59 cd/m^2, if linear illuminance (as a function of the pixel's value) is assumed and if that grey value is a $\frac{1}{32}$ step up from the full black.

That first dark tone guaranteeably perceivable would not be easy to see in all conditions, but naturally, the user interface would not need to utilize pure white screen elements either, and could present the all text and controls at some lower intensity; say, $\frac{1}{16}$ of pure white. Although such luminance level would not be a major problem in itself in brighter urban areas, the uncontrollability of the actual luminance makes it impractical even in those conditions.

The other major barrier to using smartphones or tablets directly is, that to give their answer, the test subjects would always need to look at the screen before each answer. As explained above, because of the brightness control issue, the visual

Table 7: Measured luminances in cd/m^2 for few handheld devices, with various brightness settings

		brightness setting		
	screen content	minimum	middle	maximum
HTC Explorer	all white	16	79	280
	typical user interface	11	42	179
	black	0.09	0.3	0.4
Samsung Galaxy Note 8	all white	5	270	450
	typical user interface	3.3	130	330
	black	0.01	0.25	0.55

interface would have to look different from the interfaces commonly seen in mobile applications, which would also be a possible usability issue. That is because in interface design, it is best to follow the standard interfaces, so that users don't need to learn anything new - users already spend most of their time with the other user interfaces. This is especially important for "walk up and use" applications, i.e. those that can not have long learning curve, and should be usable immediately. This input device is a natural example of a system that fits in that category. Even if the test subjects had a training period, and would then give their answers with ease, they would still need to look at the screen every time to tap the right location for their choice. These reasons would, as explained in sections 2.5.1 and 3.3, turn each answer into a diversion of attention, and would arguably affect their cognition, mood, allocation of directed attention and other affective responses - the answers would no longer be fully representative of the subject's unhindered assessment.

Therefore, the older widely known user interface, distinct physical buttons for the appraisals, should be considered better for this task. Such a button interface also satisfies the previously set requirement of being operable with gloved hands, if the buttons are reasonably positioned and if the container is suitably shaped, and provides good grip.

4.3 Design choices

All the reasons in previous sections speak in favor of a separate, specially designed input device, which should have the following technical specifications:

Clock synchronization is needed for accurately conflating location data and the input events. Mobile devices don't guarantee immediate message dispatching, but since the GPS location updates from consumer devices are usually only available once per second, a 500 ms maximum offset shall be considered tolerable. In possible future studies an even more accurate synchronization with maximum offsets of less than 10 milliseconds will facilitate reaction time testing; however such extra accuracy does not need to be maintained between trials, only for the interval from stimulus command to reaction input. In a walking speed appraisal study, the extra accuracy is not yet needed.

Status indicators are needed at least once for each device for easier pairing, but shall not be visible otherwise.

Charging connector on any mobile devices is now standardizing to Micro-USB, with many existing devices incorporating a Mini-USB connector. The wall adapters (and adapters for car cigarette lighter plugs) are available for few euros, and if the device only draws current for charging the battery, it can be charged from any USB master devices (e.g. desktop computers, laptops).

Battery size would be dictated by the average current. Review of technical specifications showed that out of the components needed, the Bluetooth radio module would have the greatest power draw when transmitting.

Battery technology for custom devices is now mainly based on the Lithium-Ion cells with 3.7 volts; they have a high energy to mass ratio, and easy and relatively low price charging circuits are available. The electronics can also easily convert their voltage to the common 3.3 V operating voltage. The only other plausible solution, rechargeable AA size batteries would be a poor choice because of their lower voltage, so that for practical operation the device would need four of such batteries, which would require considerable space inside the device.

4.3.1 Device technology

The initial set of requirements was used to find a suitable microcontroller platform for the project. Some readymade devices which could have been suitable were evaluated, but they were discarded for various reasons.

The evaluated devices and the reasons to discard them were

The Texas Instruments eZ430-Chronos is a programmable wrist watch with wireless communication, 96 segment LCD display and a 3-axis accelerometer. It could act as a hub for wireless sensors, as a standalone logger and could have a real-time connection to a PC with the included USB to wireless module. Specified battery life would have been a plus, but the device was discarded as an option, because the user interface inputs are limited to the sports watch style buttons at the sides of the device. From the buttons, there is no immediately obvious connection to the appraisals sought from the subjects, nor is there any identification related to their function in the studies. Also, the display only has the predefined LCD segments, which can show preset symbols and four digits only.

Con-Dis is a device developed and validated for collecting general perceived wellbeing state among the elderly [126]. It is a three button assessment device, which logs the data on an SD card, which is then periodically removed and the data analyzed. With an apparently extreme battery life, auditory reminders if user forgets to give feedback and three buttons for options good, neutral and bad, it would have been an otherwise very good candidate, but because it didn't have direct wireless connectivity, extending the functionality to tie the subject's appraisals to their current location and to their demographics would have required extensive work.

Wireless Bluetooth keyboards come in various sizes and have reasonably long battery life, but they have the full set of keyboard characters and other buttons, so that they can't be used without looking at them. The smallest discovered Bluetooth keyboards have the numpad portion of a normal keyboard, i.e. 16 buttons. Also, because they connect with the Human Interface Device Bluetooth profile, some mobile phones react in unpredictable ways when such keyboards are paired with the phone, i.e. they might disable the emulated on-screen keyboard, or our software could be easily closed by mistake key presses on the external keyboard. They could work in some use cases, but would not make a robust solution for different users and different devices.

Sony SmartBand SWR10 was launched in April 2014, when the initial design choices for this project were already done. It uses Bluetooth 4.0 low energy profile connection and has a battery life of several days without charging. The SmartBand has motion sensors and vibration feedback, and can be used as an input device by tapping on it one or several times, for example for skipping to the next song in the music player on the paired mobile phone. The supplied proprietary application doesn't allow recording the taps themselves, but logs the results, i.e. the song currently played at each location, and allows data retrieval of those results later. At the time of writing, it could not be a direct replacement for the device realized for this thesis for two reasons. First, it only supports newer devices with Bluetooth 4.0 support and second, it lacks a public interface for software developers; all activity and communication is only through Sony's Lifelog application.

If that software interface is later made public, it could be considered for the Living Lab crowdsourcing scenario as a possible readymade device.

When no suitable device was found, the decision was made to design and implement the simplest possible custom device, utilizing readymade construction blocks as much as possible. Building a circuit board directly around a microcontroller chip was deemed to be wasted effort, when it would not offer any benefit to the outcome. Also, several microcontroller development platforms for hobbyist are available, such as the Arduino, Teensy, BeagleBoard, Iteaduino, LilyPad and Raspberry Pi, designed with different main objectives in mind. Many of these are in fact adaptations of the Arduino platform, even if they purport to be superior for some use case.

The listed alternatives were evaluated, and rejected for different reasons. Beagleboard and RaspBerry Pi have interfaces for connecting to electronics components, but have more processing power than necessary, and because of that, high power consumption and don't work well as the starting point for handheld devices. Iteaduino has Arduino compatible boards, but their products are again more of an all-inclusive base boards which means too high current consumption, and a size unconveniently large for the project.

Finally, the choice was made between Arduino Pro Mini, LilyPad and Teensy, both latter also based on Arduino. Teensy is very similar to the finally chosen Arduino Pro Mini, but includes the USB connector for direct programming. However, with the crowdsourcing use case in mind, the price difference (2.5 times the price of Pro Mini) was deemed significant, and the Teensy requires use of special software on the developer's computer in addition to the Arduino development environment. Lilypad is a modular system designed for wearable computing, i.e. the parts could be hidden inside clothing and the wires could be run within or under the fabric. In the end, the power feed parts available specifically for Lilypad were hard to find in stock, or would have made the whole system more expensive than the more generic parts suitable for the Arduino Pro Mini.

One widely used development platform for microcontroller devices is the Arduino platform. The development of the Arduino platform was started in 2005, and it has achieved rapid success because of the ease of programming, standard USB connection, ample example programs, numerous readymade electronic building blocks for interacting with other devices and the physical world ("shields" in Arduino jargon) and an active user base. Numerous different official Arduino boards have been released, each suited to different projects, and even more board versions compatible with the development environment and the electronic properties are available from other vendors. The whole Arduino design has been released as open source, but the trademark Arduino logo may not be used in other devices.

The first Arduino boards were about the size of a playing card deck, and required an external power source, usually a mains adapter, because running off batteries was practically unfeasible. This was due to the relatively high current consumption (when compared to the current used by the microcontroller only). Later incarnations have included boards designed for low current consumption and smaller physical sizes. A version with an integrated Bluetooth module has been on the market, but it is already discontinued, and even if it could still be found from some retailers, the physical size makes it an unattractive model for this project.

Because the connectivity required from the main board of the device is limited (only some input pins, a serial connection to a Bluetooth module, and possibly an auditory or tactile feedback driver is needed), the stripped down model called Arduino Pro Mini was chosen. The board has all the microcontroller's pins as solderable holes on the outer edge of the board, and the minimum necessary auxiliary circuitry for the microcontroller, but doesn't come with a power connector or the USB–FTDI circuit. Because of this, it requires a readily available FTDI breakout board for programming, which translates the USB signal to the serial interface, available on four pins of the Pro Mini. This, however, would save in costs if more of the devices would be produced; the Pro Mini is the cheapest of the Arduino boards for sale at the time of writing, available for 10 euros from multiple outlets.

The number of electronic components was so low, that they were installed on a piece of perfboard, and the necessary connections were designed. This did affect the size considerations slightly, but made the electronic design much faster. If the device's appearance is redesigned later, the connections can be replicated in another form factor easily. The controller's operating frequency is low and connections are limited, so there was no need to further consider, for example, the cross-capacitances of the wires.

For the first version nature of the device, the enclosure was selected from commercially available plastic boxes marketed in small quantities. Some room was left on the main board for future improvements, and then the smallest possible box accommodating the electronics, the battery and the pushbuttons was chosen; the surface of the selected box was suitably rough for easy handling as is. This way the only required modifications to the box were holes for the buttons, for the USB charging connector and for the power switch. It is safe to assume that this version won't be used in torrential rain, so the enclosure can be considered suited to the use cases. Using a water protected USB connector would anyway require replacing the charging circuit with other, specialized hardware.

Unintuitively, the hardest part in the mechanical design of the device was finding suitable buttons. Many of the electronic part vendors carried only micro switch buttons, which either have a very small tactile part, very short range of travel, or a very high physical dimension inside the box, i.e. they can't be installed in a relatively thin enclosure. Thin flexible membrane switches are available, but they lack the tactile feedback. In the end suitable small spring loaded button switches, although not optimal in appearance, were found in a consumer do-it-yourself store.

4.3.2 Android software

Several decisions affect the development process in general. The design had to first consider the intended user base, their expected mobile devices and the processing power and memory capacity of those devices. The second step was to make the software components comply with the Android platform's process lifecycles, plan the common functionality of the user interface, and the messaging between the software components capable of implementing the intended workflow. Last step was to plan the data models and to implement the user interface, after the basic functionality of transmitting and receiving the appraisal button presses in the logging component was confirmed.

Although some users buy state of the art mobile phones with gigabytes of memory, there are plenty of devices with limited storage space and limited memory, both still sold and in use. Because of the possible intended use case of crowdsourcing, it was decided that the software should run satisfactorily on less powerful devices still in use. No processor intensive task was envisioned for the used workflow, and therefore the less powerful devices are not unacceptably slow.

At the time of writing the latest Android version was 4.4.4, released in June 2014, but according to Google's statistics [127], as of September 2014 still 11 % of devices were running on variants of version 2.3. Even if the last version 2.3.7 was released already in 2011, such devices were still sold in the end of 2013. Because of these reasons, the software is designed to work on devices running Android 2.3.3 or later, also known as API level 10.

If volunteer subjects are to install the software on their own mobile phones, it is necessary to support a wide variety of devices. Out of the three major operating systems in use, only Android allows developing and installing one's own software without costs, both locally and through the public app store. Also, should there be a need for a single party to purchase several handsets for such concurrent multi subject research task in the future, the cheapest devices should not be unacquirable because of their price.

With this in mind, the size of the compiled software should be kept small, and the code functionality should be based on the Android version with a reasonably high user base. At the time of starting this project, it was Android version 2.2, i.e. API level 8. Later API versions do not contain features imperative for realizing the goals of this project. Version 4.0.3 (API level 15) does contain improved Bluetooth functionality, which could become handy if the system is later developed further.

Bluetooth 4.0 Low Energy modules will also soon be inexpensive, and have a very low current draw. This could be used to improve the device, but support for the low energy connection profile requires Android version 4.3 (API level 18) on the mobile device.

Android ecosystem has many active open source coders, who have published usable packages and classes that could be included in other projects. It is, however, easy to increase the size of the application considerably by including packages that contain not only the needed code, but which also contain even megabytes worth of superfluous code then left unused.

Even the often included package, the Android Support Library, can be bigger than the application would be in itself. The support library adds some of the functionality introduced in later versions of the operating system for software running on older devices, such as common user interface elements. This comes at the expense of a significant size increase, though. A minimal Android application is only some tens of kilobytes in size, but with the basic functionality, in the early stages of development, the size of the installation package was about 280 kilobytes. Depending on the selected version, the Support Library can increase the size of the application from 300 kilobytes to over one megabyte. Because of this, it was decided that including it should be avoided, unless absolutely necessary. The final size of the software developed in this thesis ended up at 508 kilobytes, which is to say this approach succeeded in keeping the software size small.

In the end one external library was required for showing the map background and the recorded appraisals overlaid on said map. The library of choice was Osmdroid, an open source map library replicating the functionality of Google's map library.

In stock condition, Android devices only allow software to be installed from the Google Play Store, or the device manufacturer's own application store. However, if the user changes an option in the device settings, any compiled Android software can be installed from any source once the package has been transferred to the device. On devices running 4.2 or newer, that option is initially hidden and only visible after a semisecret tap sequence is performed, for example on devices running Android version 4.2, user has to select the About phone option in preferences, and tap the menu item Build number seven times, after which the Developer options menu is visible in the previous menu level. In essence, this is something usually only application developers do with their own code.

Because of these unnecessary steps to install self compiled applications outside the application store, it was decided that the software should conform to Google Play Store requirements, so that any party interested in reusing the code could likewise publish their version in there and have it easily installed on the necessary devices. This would also be a necessity, if the device would be produced in larger numbers to facilitate a true Living Lab crowdsourcing setup.

This choice has at least one disadvantage when compared to the option of not conforming to the Google Play Store requirements: the API documentation states that software must not turn on the Bluetooth adapter on the device without user interaction, if the user has disabled it. Even if the only reason to use the logging software is to connect to the input device over Bluetooth, because of this, the software has to ask for permission before turning on the Bluetooth adapter, if it is not enabled already. The user can, however, give a permission to do so in the future without asking.

4.4 Implementation overview

4.4.1 Device

The device is a based on a *Arduino Pro Mini 3.3 V 8 MHz* board, which is a version of the Arduino development platform stripped to essentials, but ready for programming almost as is. It doesn't come with any headers soldered, but any necessary headers can be soldered in easily.

The Pro Mini doesn't have a ready installed power connector, but can accept a ready regulated 3.3 V feed. It also has an internal regulator, which can accept anything between 3.35 V and 12 V. A suitable power source is a 3.7 V Li-Ion battery, which are available in various physical sizes and with diverse capacities. Ready-made charging circuits are also available, which have a Mini-USB connector for power input, and internal connectors for the device. This way the designer does not have to concern himself with the power delivery and loading connectors.

For programming the microcontroller, four pins on the Pro Mini are used for a serial interface, with readily available USB–FTDI adapter cables and FTDI breakout boards, which can reset the Arduino board automatically to start the programming.

The code for Arduino devices is developed in the Arduino programming environment, which is based on two open source projects: the Processing environment and Wiring framework. The code is written in a C language variant, which abstracts away many of the low level details.

The third building block for the device is the Bluetooth module, which appears as a transparent serial connection on the Arduino. The hardware only requires logic level transmit in and out pins, and operating voltage pins. The available modules vary in their functions and in the amount of commands required from the Arduino to the module before a connection is set up properly, and established. After reviewing specifications and published user experiences, a product by the name *Seeedstudio HC-06 Brick* was chosen. If there is only one powered device with that module in the wireless range, no setup commands are required, and only when the module is considered for pairing with the mobile phone, it has to be identified by its name. The device has a blinking status led when it is not connected to any other Bluetooth device, and it turns to a steady indicator light when a connection is established.

The device was constructed in a box, which turned out to be partially translucent so that the indicator lights of the Bluetooth module and that of the charging circuit shimmer through the enclosure's wall at road lighting luminance levels. This eliminated one step in the construction, when no extra holes, parts or code were needed for the indicators.

The location of the input buttons was selected such that the user could identify them with touch. They were placed almost at the other end of the enclosure, so that the device could not be held upside down by mistake, but near enough to both long edges of the device that the device could be used in either hand. This allows the test subjects to give their feedback totally without looking at the device, so they can go about their movement without changing the allocation of visual attention.

If other different but suitable and more durable buttons are found and acquired, the enclosure could be replaced with another variant easily: there is a three pin connector on the internal perfboard for the buttons and for the common ground for the buttons. With the ground level intentionally in the middle pin, plugging in the connector the wrong way can't break anything; only the mapping from buttons to logged values is reversed, but the values can be set in the software.

The second connector on the circuit board is for the two lines from the power feed, again with holes for three male pins, with the ground level in the middle. If the connector is plugged in backwards, the device just doesn't get any operating voltage. The last, third connector has four angled header pins for the Bluetooth module, which is in this instance connected with a short readymade cable. This allowed locating the module more freely within the enclosure. Some space was left on the circuit board, for future connectors for connecting more of the Arduino's input pins to different input methods' outputs. The programming interface required for possible future improvements is available on angled header pins soldered directly to the Arduino. Accessing the programming interface only requires the removal of the box's cover, and disconnecting the voltage pin to Bluetooth module. The Bluetooth must be disconnected, because the module unavoidably connects with the same serial transmit in/out pins as the programming interface.

4.4.2 Android software

Additional features that needed to be considered in the normal operation of the software were numerous. The software has to be prepared to handle, for example, a disconnected Bluetooth connection, an unexpectedly removed external mass storage (SD-card), a lack of GPS signal at startup or during recording, varying results from mailing the resulting log files and differences between device manufacturers' implementations of the Bluetooth connection establishing functions. Also, the Android operating system can stop or restart any activity at any time - such as when the device is turned sideways or back upright - and the software must handle these restarts, even if the operating system does provide some help with that.

The application contains five different screens, called Activities in the Android system, and a sixth one for the common settings. The recording and Bluetooth connection are managed and performed in a service; in Android a service is a component that does not have any activity attached to it. Services can, however, set themselves as foreground tasks, which mostly prevents the system from stopping the service until it releases the foreground status, or the device is so low in memory that system services require to do so - in which case, with this application, the recording would have to be considered lost anyway. Foreground services must be something that the user is actively aware of, and must provide a notification for the system's status bar so that the user can see if the service remains running.

All of the activities extend a custom base activity class. This was done to ensure all activities register their message handler with the running service in the same manner. Any activities, except for the activity that started the service, don't initially and automatically have any means to send messages to the service, or to receive messages from the service. Automatically registering the current activity with the service also allows the service to ask the current activity to, for example, modify the user interface, which can't be done from the service's own thread. This is used to indicate the button presses of the appraisal device on the screen, and to show the dialog to ask for permission to turn on Bluetooth.

The simple so called root view is visible at startup and between subjects. The only user actions possible are starting a new test subject, showing a list of logs, i.e. previously recorded subjects, and opening the settings from the menu. Hidden in the menu is also the possibility to shut down the invisible parts of the application, just in case it had crashed in an unexpected way. For example listening for GPS location updates can consume the device's battery rapidly, so it is good for the user to have a remedy in such unlikely, but possible cases. The root view also shows the status of the device connection, i.e. whether the system is ready to record almost instantly, or whether the connection and current location need to be established first.

When a new subject is started, the next view asks for the subject's demographics. At this time, only gender and age are asked, and both can be left undetermined. Extending this view with more questions would be straightforward. These answers are only stored as tags, i.e. freeform key and value pairs, until they are saved in the log and analysis is left to desktop applications of researcher's choice. The screen also shows appraisal device button presses for 0.5 seconds each as text lines at the top of the screen, so the correct functioning can be verified before the subject is sent en route.

When the software has connected with the Bluetooth input device and has a GPS location, the button to start recording is enabled and the user can continue to the actual recording. In this third view, the recording in progress screen, the only options are to edit the demographics again, and to save the log when finished. Because the Android device is typically put in the subject's pocket at this point, and the screen turned off, the use of the back button is prevented to minimize any mistake actions. While the screen is on, the small map view will follow the user's location and button presses are again shown for 0.5 seconds as texts under the map, to confirm the system still works as intended.

When the data is complete, the researcher or subject presses the Finish recording button, and is taken to the last map view, which shows the subject's route and their inputs on the map. Now the researcher can visually identify areas of interest on the map (be it areas with many positive or negative presses, or with little inputs), and add a comment to any selected recorded location. Appraisal points are shown with red and green dots, whereas other recorded locations show up as black, smaller dots. The selected point is highlighted, and the Add comment button is only visible when a point is selected (shown in Figure 1). Each entered comment is saved immediately, but they can be always re-edited. Tapping far away from any points deselects any points. Once the necessary annotations have been written, the recording is finished.

At this point, the log is emailed straight away, if set so in the preferences, to a previously set address. The user is then taken back to the root view, and the Bluetooth connection and GPS functionality is turned off to save battery.

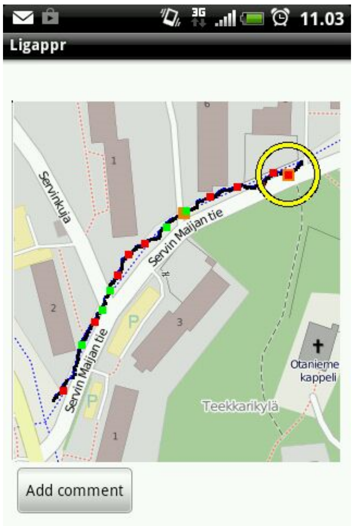

Figure 1: Screenshot of the commenting phase in actual size on one device, with an already commented appraisal selected for commenting

It is however probable that the logs are easier to analyse if they can be sent combined so that the answers of several subjects are in one single file. The root view gives access to a list of previous logs, where several logs can be selected simultaneously (either by picking individual logs, or by choosing the option to add to the selection all logs from one date), and all of them then sent in one message. The list also allows deleting files, and showing them on the map individually. When a log is shown again on the map, comments can again be added to the points. This can be a crucial feature for example if the recording was ended too early by mistake.

5 Validation of device functionality

To assess whether the device was worth the effort, the functionality, usability and possible benefits for current research goals needs to be addressed. A button is just a button, until we know it does the expected.

5.1 Step towards a Living Lab

One of the ultimate goals for the research device depicted in this thesis is the applicability to a Living Lab environment. Therefore, the method developed for this thesis can be reflected against the LL model.

Tang proposed a LL process model (Figure 2) in his dissertation [53, p. 36]. Before listing the process model phases, it must be noted that in a true LL, the phases can happen concurrently; the users in the core of the model facilitate knowledge transfer between these phases in an asynchronous manner.

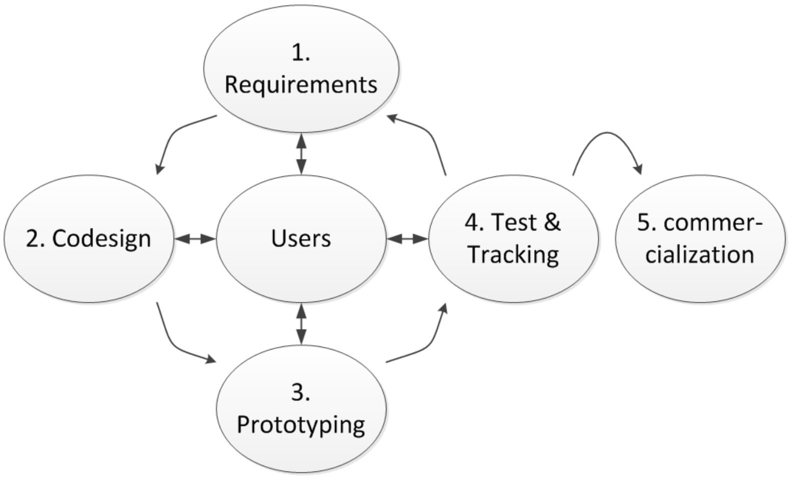

Figure 2: Living Lab process model, according to Tang [53]

In the first stage the users and contexts are preliminary identified, along with the issues to be solved. These steps have been researched outside of this thesis. Already in the next stage, however, the designers codesign some LL services and products, with the stakeholders and the users. In the third stage the focus is in prototyping actual services and products related to the context of that particular Living Lab, with the users. The fourth stage consists of tracking and evaluating the users' performance and interactions with the prototypes in their real-world context. If the feedback is positive, a commercial opportunity might be identified, and the new solution can be commercialized.

It is trivially evident that the device and software are not innovations related to the actual lighting or to the lighting control in any area, but they are tools for tracking and evaluating the users' performance and interactions with, say, service innovation prototypes. As such, the results of this thesis are a step towards successfully implementing and conducting the fourth stage of the Living Lab process model. Given the validity of the produced data and usability, reported in the next chapters, the method can also be considered as better than pen and paper questionnaires, and a step towards implementing and conducting that stage of the LL process model better.

5.2 Validation of data output

The data logs produced by the system need to be analyzed after the study is completed. The validity of the data should be first defined, and then verified, to be sure that the system output conveys something meaningful about the environment. Because this kind of appraisal collection has not been performed before in the context of lighting, there are no established criteria we could use. Rather, we have to rely on criteria derived from the research question and from the goals set in the previous chapter, and on post-hoc analysis of one test night's data.

In line with the goals of product development, set in chapter 4.1, the log files contain lines, with each line containing a time stamp and coordinates of the device at that time. Depending on the software settings, the test subject's demographics are either on each line, or at the start or at the end of the file. An example excerpt of the output log's beginning is given:

```
eventAt,Latitude,Longitude,button,comment
2014-10-27 17:32:56, 60.18912462,24.83466343, , , Age=0,Gender=Female
2014-10-27 17:32:57, 60.18912462,24.83466343, Button 2, , Age=0,Gende
r=Female
2014-10-27 17:32:57, 60.1891312,24.8346791, , , Age=0,Gender=Female
2014-10-27 17:32:58, 60.18913923,24.83469431, , , Age=0,Gender=Female
```

Later in the log file is also a commented appraisal row:

```
2014-10-27 17:34:22, 60.18992999,24.83585201, Button 1, light was on.
open space. , Age=0,Gender=Female
```

These excerpts show, that the log contains the required values, and the system has recorded the comments. The demographics were written on each line, because the application settings were set so; Age=0 denotes the fact that no age was given. By coincidence, the comment shown here happens to coincide with the findings of prior studies, that people might prefer areas with higher prospect and lower entrapment, also at dark times, and also that giving a general assessment of the lighting draws from various environmental attributes.

Hence, the goal specific to the data output has been reached, but also the value of the data for analysing the area's lighting has to be verified. To this end, the data from the user test was analyzed; the procedure of the test is described in the next chapter 5.3 on usability validation.

Because the users' appraisals can have any coordinates, which don't directly relate to a single luminaire, there must be a way to somehow derive meaningful information from the distribution of the appraisals. Common methods are heat maps and pattern recognition, to identify how to divide the samples into clusters of similar values.

The five test subjects' logs were concatenated in an ordinary spreadsheet software, and all data points' coordinates were mapped to a 10 by 10 grid, effectively dividing the area covered by the points into 100 bins. The calculations were then repeated with only the points with appraisals. This allows us to generate a surface chart, which depicts the distribution of the positive and negative appraisals. Scaling the data was not yet considered in more detail, and the figure 3 represent the difference in the number of positive and negative button presses in each bin, divided by the sum of button presses in that bin. This way a bin with only positive appraisals gets the value of one, and a bin with only negative appraisals gets a -1. A map image of the area is overlaid with transparency, for easier identification of the indicated locations.

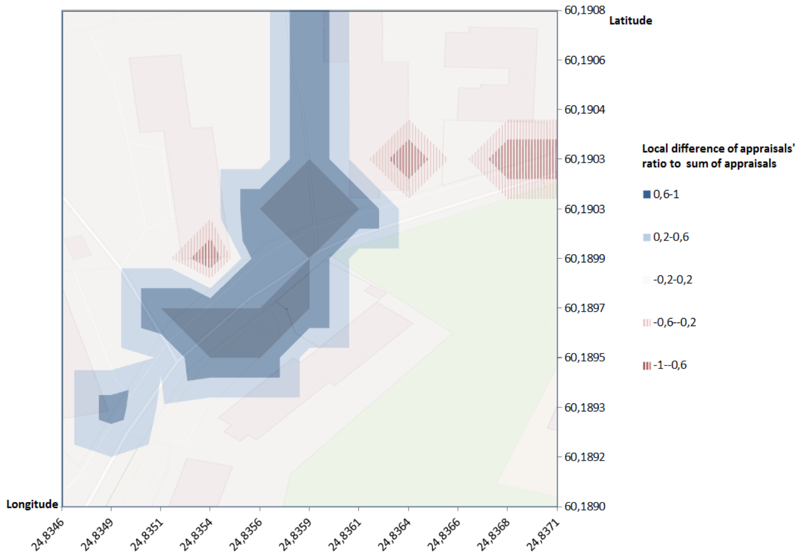

Figure 3: A heat map from test users' appraisal

Then, figure 4 contains a histogram plot of the calculated values with these parameters. The resulting figure is intuitive: given the limited number of subjects, and the fact that the bounding box of the area encompasses also areas where none of the subjects walked, most of the bins were empty, and the zero value dominates the histogram. However, one interesting notion can be deduced: there were more bins with only positive appraisals than there were bins with mixed appraisals (with values near zero). This can be viewed as indicative of the fact that people have liked some spots unanimously, but the limited data is not worth anything more than a hypothesis for future research.

The plot in figure 3 revealed that one test subject had held depressed the positive appraisal button for the duration of his walk, which resulted in a significant bias in the results, seen as a large uniform area of "good" appraisals. He had a personally valid reason for the response, given that he came from a country where

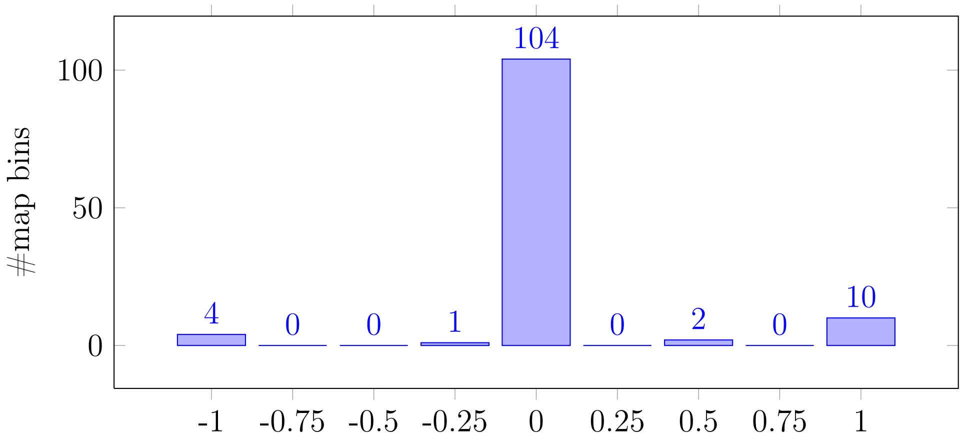

Figure 4: Distribution of the scaled map bin values

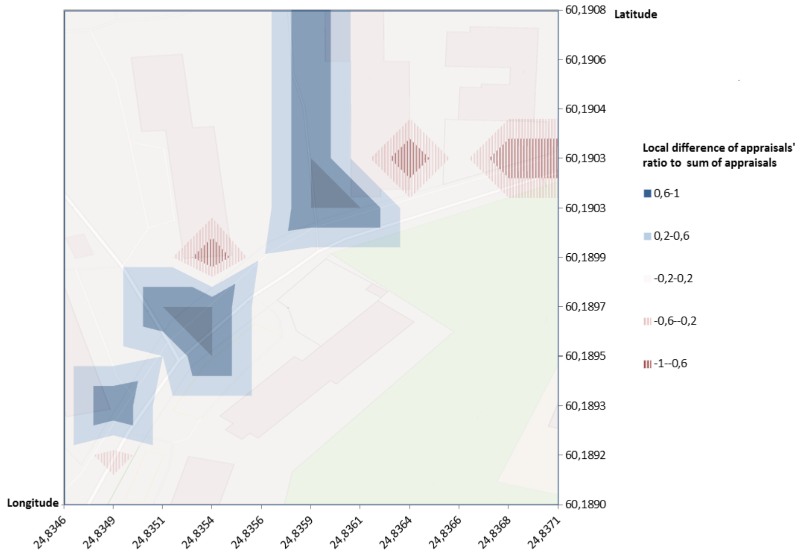

Figure 5: A heat map with one test user's appraisal removed

street lighting was exceptionally poor, if not mostly nonexistent. Had there been more test subjects, such anomalies could have been detected by comparing the distributions of the appraisals within subjects, and the distribution of time intervals between appraisals. As a remedy, the data for that subject was not considered in any further calculations, and the heat map of the remaining subjects' opinions is shown in figure 5. For visual comparison, each subject's data was separately quantified with the same method, and the results for one subject are shown in figure 6.

The optimal geographical size of each bin depends at least on the area, route, number of subjects and test setup: if the area is divided into too small areas, each bin might only have one appraisal by one subject, and if the bins are too large, the differences between luminaires can't be identified as the average and the distribution of appraisals in each bin approaches the general results of the whole area. It would also be interesting to further study what would be the most relevant indicator value

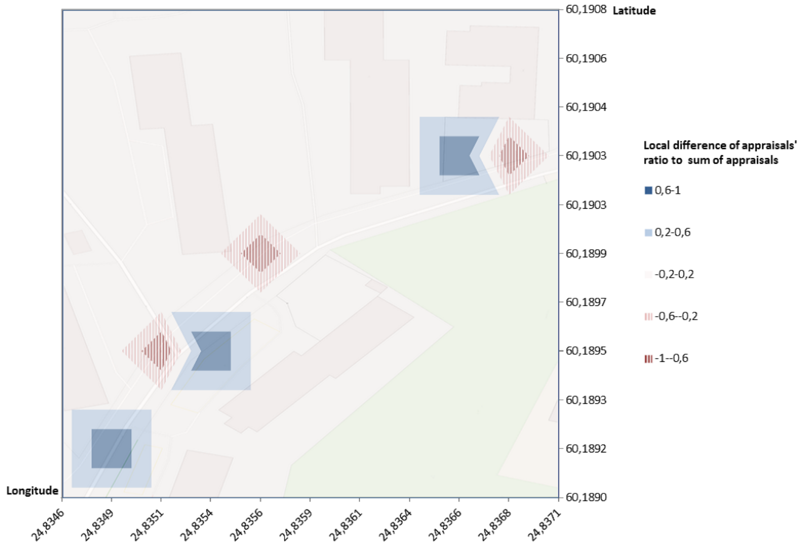

Figure 6: A quantified map from one test user's appraisal

calculated from the appraisals in each bin, but such an analysis would require a much larger sample.

The raw appraisal data points could also be analyzed in other meaningful ways. Data that can be presented on two or several axes are often classified and explored in pattern recognition with tested algorithms. The algorithms work by extracting features from the known values, and then using those features to train a prediction formula or system. For this data, no intersample relations were considered as features, but examples of such could include the elapsed time from previous appraisal, and the direction of movement calculated from the location of the user one second prior to the appraisal. Increasing the number of features increases the dimensions in the description vectors, and with huge datasets, the analysis could even take considerable time.

This appraisal data could be easily classified with the k-nearest neighbours algorithm (k-NN). Increasing the number of extracted features would increase the amount of algorithms to consider, from support vector machines to neural networks. For the purpose of only showing that the data can be used in this manner, and with feature vectors of mere three dimensions and given the small number of samples, the k-NN should fare no worse than more laborous methods. In the k-NN the algorithm calculates the distances between all samples, and assigns a new value to each sample by taking the most common value from the nearest k samples, where k is an integer, and usually odd to ensure either classification has the local majority.

Selecting the optimal value of k is not trivial, but the usually suggested method is bootstrapping and repeated random sub-sampling validation for error rate minimization. In short, the original data is divided to training samples and validation samples at random. Then for each validation sample, the classification by nearest neighbors is calculated only considering the training samples, and the error rate of this classification vs. validation sample is calculated. This is repeated with random subdivision of the original data for tens of iterations, until the minimum error rate for some k is found. If the value of k is too high, the algorithm starts to loose local variations, but if the value is too low, the result is overlearning, i.e. the classifier only replicates the original.

The algorithm was implemented in spreadsheet software, and the results show that some of the locations identified above can also be identified with this method. The eastern end of the test routes has several negative samples after the k-NN is applied, for different values of k (Figure 7). Repeated random sub-sampling validation with 30 iterations gives the results of figure 8, which can be used to say that given this sample data, if the value of k is 3, the produced classification is most likely the best that can be achieved with the k-NN method. With a more spatially dispersed and bigger dataset the result could be different, so the analysis should be repeated for any appraisal study. Other algorithms for developing reliable sample classification functions should also be examined if the data collection method is used for the analysis of a real location's lighting.

(a) k = 3 (b) k = 5

Figure 7: K-nearest neighbor classification of appraisals. Outlined circles indicate appraisals as "poor".

Given the locations of the appraisals, and the locations of the luminaires, one could also approximate which luminaires were visible to the test subject at that instant, even if the data does not directly record the users gaze directions. Because the location can be recorded between button presses once every second, the direction of travel at that point is known with some accuracy. In areas with good open view of the sky, the number of visible GPS satellites can be high and the sources of interference insignificant for the between samples location accuracy. In such conditions, the relative error between consecutive recorded locations is small, even if the actual location is easily 5 to 15 meters away from the indicated coordinates. The error is, however, often smaller than that, but the absolute value of the error can not be reliably predicted from the values reported by the GPS receiver. This

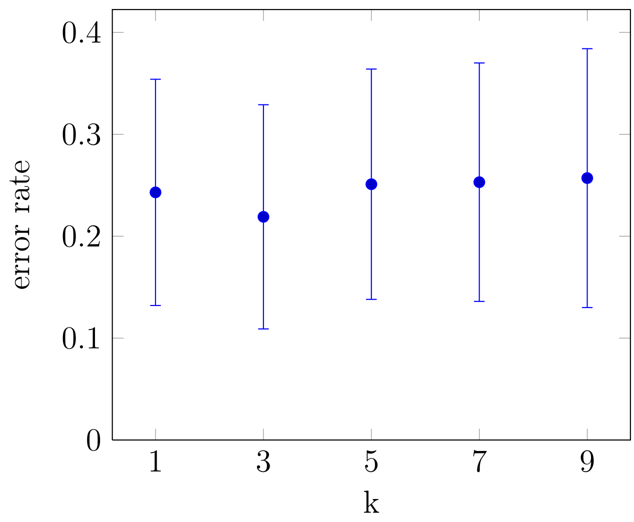

Figure 8: Average validation subsample error rate with different k, with 30 iterations

means, that the general direction at each point is known, outside of areas where the GPS receiver has difficulties obtaining accurate signals. Such difficult areas would include for example densely built urban streets with apartment buildings touching the sidewalk.

To check the feasibility of luminaire visibility analysis, the expected gaze direction was calculated for one test users' one indicated poor lighting location, where he commented having experienced glare. The direction was calculated from the location of the button press (the small circle) and the previous location point, and also from the point taken as the average of three previous locations (the small triangle). These predicted lines of sight are shown in figure 9 as directed arrows and assume pedestrians generally look straight ahead. The other circles represent luminaires that were possibly visible, and the measured angles in the figure indicate the horizontal eccentricity of a later identified luminaire causing major glare, for both of these approximated lines of sight. The numbers are not important, but indicate that the same set of luminaires would have been assumed to be visible, with both predicted gaze directions.

If the system is used on a set test route with different installations for the subjects to compare, there's one conclusion that can be derived from the accuracy of GPS positions. If the distance between different installations is at least 30 meters, i.e., twice the generally accepted 15 meter error probability, then the expected mean error is smaller than the distance between luminaires. This means that the appraisals can be tied to the correct luminaire without manual adjustments. Smaller distances can be used, but they might require manual verification of the appraisal log's location data for apparent deviations.

All in all, the data was shown to contain the intended information, and the information was shown here to yield knowledge of the spatial variations in the test subjects' appraisals.

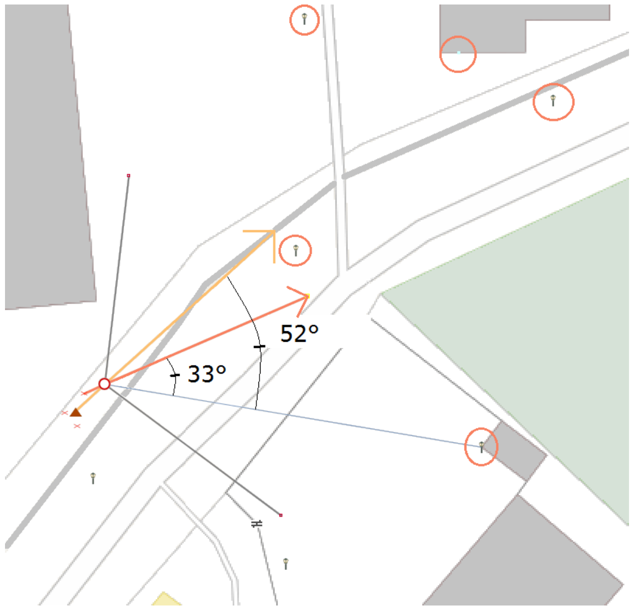

Figure 9: Feasibility study of visible luminaire analysis for an indicated poor lighting location. Identified glare source was plausibly on the right side of the visual field

5.3 Validation of usability

It is not sufficient to claim that the device produces usable data when a trained developer uses it. The usability of the device depends on the user interface accomplishing the ease of use and efficiency for the purpose, having good learnability and user satisfaction, with real users.

5.3.1 User tests

A usability test was conducted on a Monday evening, starting 20 minutes after sunset and 40 minutes after the street lights had turned on. The weather was dry, temperature was 10 °C and there was no snow cover. This was to ensure that the lighting context was comparable to the actual user tests that would be conducted with the device in the future.

Although more test subjects can give more comprehensive results, it is generally considered better to test device user interface iterations more often with a smaller sample, rather than have a larger sample when the device can no longer be modified, or when only minor changes are possible. In general, even three test users can find the majority of the important usability problems with an interface, even if somewhat higher numbers spread across the development iterations are recommended for optimal effectiveness [128, p. 212], and based on experiences from decades of user tests, on average the pay-off ratio between the monetary benefits and the cost of running one more test user is at its highest with just three users. [65, p. 173]

The five volunteers (three male and two female; two with background in lighting engineering) walked road sections 160 to 240 meters long on routes they were about to use in their normal activities. The first 120 meters of all routes were identical, because there were no intersections in that part of the pedestrian footway. They used the device to indicate where they felt the lighting was poor, or good, whenever they would be confident to have an opinion. At the end of their routes, they filled in a one page questionnaire about the usability of the device. The two researchers also answered an almost identical questionnaire, with relevant questions about their user experience. Later, one subject performed the task while driving a passenger car for 4.5 km; the device was in a cup holder next to the gear shift lever, so the driver could operate the device without holding it in his hand, nor was he required to move his hand far from the location where it is usually positioned, nor to hold it in a raised position.

The test subjects indicated that on average they found it easy to give their answers, with only one subject somewhat disagreeing (in table 8). The system was pleasant to use, and most subjects felt that when using the device they did not have to concentrate on the device, which means the system does not seem to interfere with the directed attention of the subjects.

The open ended questions provide more detailed and more useful depictions of their experiences. Two volunteers had comments about the device's physical characteristics, whereas three subjects indicated that at first they had trouble understanding the task, i.e. when should they press the buttons, or in other words what property of the lighting they could be assessing, but had no other difficulties. This can be an indication that the basic idea of using the device depicted in this thesis is possible for lighting appraisals, but more emphasis should be placed on the actual study question and especially how it is presented to the test subjects. The system does allow the subjects to practice before the actual recording is started, but in this user test such a trial section's effect on the user experience was not tested.

The indicated problems with the construction were that one subject felt that it was a bit awkward to hold the device in his hand, and one subject would have liked to have better signs and colours on the buttons better indicating positive and negative.

Table 8: Number of user's agreeing to a subset of statements asked

		It was easy to give my answer	The system was pleasant to use	Device required concentration
5	Totally agree	2	1	
4	Somewhat agree	2	4	1
3	Neither agree or disagree	1	1	
2	Somewhat disagree	1		2
1	Totally disagree			3
	Average agreement score	3.8 Almost somewhat agree	4.0 Somewhat agree	1.8 More than somewhat disagree

Two subjects indicated that they envisioned this would help make the road environment safer.

Looking at the answers by the researchers who tried to use the system, several opportunities for improvement were found. The most significant point was that recording the open ended comments given by the test subjects at the end of the test was somewhat slow and unnerving, as the text had to be written on the mobile phone's small onscreen keyboard. They also noticed that if the weather were any colder, entering the comments would feel uncomfortable after many subjects.

The first of these challenges can be addressed to some extent by using a mobile phone or tablet computer with a bigger screen and better onscreen keyboard software. When this system is used on a set test route, the researcher can choose more freely (as opposed to the LL scenario) from commercially available mobile phones, or use their own device which they are accustomed to. The first alternative out of the two solutions considered, using a video camera affixed to the test subjects head and having them explain their feelings as they walk, would make the whole device unnecessary, but has the burden of data input when the videos are encoded for analysis. The second alternative, recording the comments as sound files within the software, had only been considered a few days before the user test session and could not yet be implemented, but that could help make the recording of their opinions easier. Recording audio files is claimed to be straightforward in Android developer documentation, so that could be one of the first improvements to consider if this system is developed further.

The second most important observation from the researchers' point of view was that the startup screen's status display wasn't very intuitive when the mobile phone was connecting to the input device; once the connection and GPS location was established, it was immediately clear that it would indicate the state correctly when the recording can start.

On the positive remarks, selecting the points where the comments would be asked was however easy, since the path taken and the appraisals were presented on the map. This was true both for those seeking comments for each appraisal point, and the other researcher only asking questions about the (visually selected) places which had several positive or negative appraisals.

The user test in itself and the comments given by the test subjects did provide some confirmation for other aspects considered in this thesis. Many pedestrians do consider that lighting increases safety, glare is one the biggest factors in assessing the overall goodness of lighting and the affective evaluation of lighting does not depend solely on the measurable characteristics of the lighting (such as distribution, luminances and CCT), but also on the physical characteristics of the environment, such as openness. Likewise, the quantitative appraisals of lighting do seem to provide solid information of the locations, but the qualitative comments given by the subjects reveal more insight and can tell us something about the phenomena affecting their appraisals.

6 Discussion

6.1 From literature to device requirements

Requirements engineering is a separate field of study that tries to model how the requirements for a computer based system are discovered, documented and managed in an effective way. An ad-hoc method of coming up with the system requirements from vague descriptions of planned use cases is likely to lead to missed requirements and implementation challenges later in the development. Even for a simple system as the one developed in this thesis, the requirements or goals would have been better and more importantly, more to the point, if they had been scrutinized in more detail already at the beginning of the process.

All aspects of development process can be improved and modelled better, but whether the application of more laborous methods pays off in the end is a relevant question at each step of the process. Even the best available methods may dictate more work than can ever be saved with their application, if the target system is simple enough.

However, the goals and requirements set for the developed system can be positioned on two axes: first, whether they relate to and support the test subjects in future tests in assessing the lighting, and second, whether they relate to the technical construction details.

With that in mind, the requirements relevant for the studies of lighting appraisals were relevant. These are individually discussed below:

- Data points shall be tied to a location. In view of the reviewed studies and concepts, the context matters when people's perceptions are studied. The literature review includes several studies showing that affective responses to lighting are different in different contexts, both in different locations and when the same location is altered. This requirement provides essential data for analysis.

- The fact that the test subjects don't need to wait for the device to operate it, was found in the user test to be a good idea: with the simple interface, subjects started using the buttons almost immediately, so that if such action would have lead to an error situation, several of the test users would have had problems with the device.

  One initial premise for the whole development was, that giving the appraisals at the very instant of the experience the answers would better reflect the actual experience, as opposed to the already altered memory of an experience. The literature review found support for this idea, that instant appraisals are generally better than those collected later.

- The possibility to use the device while driving vehicles is a good idea, but actual research would need to first investigate how and where to affix the device for best operability and for least inconvenience and driver distraction.

- In the user test with a driver, the missing option to remind them would have yielded more appraisals. Movement requires variable attention, and when allocation of attention is changed with time, the device and the task of giving appraisals can be forgotten for a while. This can be viewed as good, though, as even the lack of appraisals in some area or location can tell us something about the responses of the test subject.

- The goal of gathering subjects' qualitative comments does help in analysing their motivations and reasons, beyond the pure affective responses. Without these, the "why" of the possibly identified locations would only repeat the accepted knowledge; in the user test, only the glaring light source would have been identified with the expert inspection of the pinpointed location.

The remaining two requirements relate to the construction details:

- The battery life was not a limiting factor in the test. If a smaller battery had been chosen, it would have ran out if more subjects were tested in a real study setup. The physical size of the battery is not the limiting factor if redesigning the enclosure, so the selected requirement does not have bad or negative consequences on other attributes of the system. If the number of test subjecs can be increased, as suggested in chapter 3.11.2 based on prior research, a small capacity battery would become a limiting factor for the number of one night's test subjects. Based on this, fulfilling this requirement is a crucial technical feature for the device.

- The fact that the device does not yet have any label of it's identity printed on it would be a problem if more of the devices were produced. Although only the one device with power switched on can be visible to other Bluetooth devices, the status indicators are only dimly visible through the plastic enclosure and not at all in office conditions. Once the Bluetooth devices have paired, it does not matter if they have the same identifying name; on the link layer level, the connection is based on the 48 bit device address of the adapter. This requirement was therefore a good one, but of limited value to the proper functioning of the device, and not related to the research objectives in view of the reviewed research.

6.2 The device in view of requirements

The important functionality goals were reached. The device can be used for collecting user appraisals with location data, without users looking at the device, and the locations can be commented at the end of the test route.

The usability was evaluated, and was deemed a success. Even if opportunities for improvement were identified in the user test, no critical usability errors were found.

Numerous classification systems for potential and real usability problems have been proposed, but they assess the probability of encountering the error, and the impact it would have on accomplishing the goals, from preventing the task to minor inconvenience. Critical problems are those that are encountered frequently, occur in a fundamental use case action path, and are difficult or impossible to circumvent. They can prevent the completion of the task altogether. The lowest level problems are distractions that are easy to overcome, either rare or encountered only on some non-fundamental action paths, and the problems falling between these classifications could be divided into two or more groups. Based on the constantly recurring nature of the identified room for improvement in entering the comments of locations, improving that aspect of the user experience should get the highest rating out of those identified in the user test.

Out of the goals, functionality requirements and technical requirements set for the device before the development started, described in chapter 4.1, two feature goals were not yet met:

- The device does not yet have the option to remind the subjects by vibration or sound if they have not given any feedback with the device in some time. There is working code in the device to respond to commands sent from the Android software, but without any ways of defining when to remind the user, the effort required for adding the output method to the device was spent elsewhere. In a real LL environment, this functionality would most likely improve the amount of crowdsourced data, and would help avoid gaps between locations, when users could be reminded to use the device, for example when they enter areas without previous appraisals.

  For a set test route, the user test did show that subjects seem to indicate their appraisals at regular intervals anyway, so this lack does not prevent the usage. On a longer route when driving a vehicle, there was a gap in the appraisals, so it seems this should be implemented if subjects are to be sent to wander around a site in a more freeform manner.

- The device doesn't have an identifier both printed on it and available wirelessly, but rather uses the default name given by the Bluetooth module manufacturer. This is because it was deemed important only if several of these devices would be in use, and other features needed to be finished within the time allotted. The Android software already has the option to search for BT devices with any given name, but changing the name was not implemented on the input device. The other devices can be turned off for the duration of BT pairing, which only needs to be done once for each device pair, so this lack does not prevent usage even in a LL case. That functionality should, though, improve the user experience.

One technical requirement set in chapter 4.2 was only partially met: although the data connection is used to download the background map for the appraisals, and the map images are cached on the device, the system can be used without a data connection. If there is no data connection and the area wasn't cached prior to the research event, the background is just grey. For a set installation test route, the researcher most likely has visited the site beforehand and the map is available after the first test. In a Living Lab use case, the participating citizens most likely do have and need to have a data connection enabled device for other functions of the LL. Circumventing this, i.e. to have maps even when a mobile data connection hasn't ever been available, the map display library would need to be modified.

The remaining goals and requirements were completely met. Looking at the device, using the input method demonstrably only requires the new device and an Android mobile phone or tablet. Test subjects' trial runs succeeded with gloved hands, in the cold; researchers couldn't enter the comments with gloves, but this could be circumvented with a touch screen stylus pen if necessary. Based on the consumption, battery capacity and experiences during the software development, the battery does last longer than the researchers in one evening. The one subject test showed that the device can also be used in vehicles and while driving vehicles, and the system facilitates the recording of qualitative affective comments about the locations.

In closing, the more important goals, requirements and usability measures were met. For a real LL test with more devices used in the subjects' daily lives, the identifier and reminder shortcomings would preferably be first addressed. Improving the design of the container could incorporate better, waterproof charging connector and waterproof buttons.

6.3 Known limitations

The Android software was designed for API level 10, so at some point in the future new Android devices may no longer support it, although current devices can run even API level 1 software. There are some small sections in the code that use API methods that have been deprecated in later API versions, but at the moment of writing they could be trivially replaced with the newer methods. Those newer methods were not available on API 10 devices, and could not be used to allow such devices to run it.

The device records the appraisal logs to one comma separated values file (CSV) for every test subject, and they can be sent to an email address in a batch, or one file at a time. Concatenating the logs to one file was not implemented, because the desktop computer software that would be used for analysing the data is not known, and hence the correct file format is not known, either. Rewriting the logs to another format could be implemented in the software. Another option would be to extend the previous logs listing activity to allow statistical analyses to be performed on the combination of selected logs, and the results could be shown directly over the map. While testing, it was noticed that that same log listing could be improved with more detailed descriptions of the log files in the list; now only the date and the number of recorded rows is shown.

7 Summary

This thesis has reviewed prior research reports of lighting studies conducted with users, and developed and validated the input method for studying users' opinions of outdoor lighting.

The goal and aim of literature review was not to consider the validity of their findings, but to document and to explore their methods, and the concepts that have been studied in relation to lighting. Based on those studies' results and the models of environmental psychology for subjective assessments of the environment – as they were introduced in those studies – it was considered evident, that the users' appraisals of the lighting is first and foremost predicted by the three measurable qualities of lighting: brightness, lack of glare and lack of darker areas (one aspect of evenness). After those measures, the subjective attribute that is usually called naturalness has had the largest impact on the volunteers' appraisals.

Analysing the meaning of naturalness, as it has been applied in prior work, the cues of naturalness were now claimed to be the extent to which the lighting reproduces daylike lighting conditions; the environment and the people are perceived as naturally lit. Obviously there is less light available at night, but people should be able to ignore the absolute level and to judge the resemblance; "if there was just more light, could I then see this same view?"

Expressed in lighting terminology, I would hypothesise (based on the findings of prior studies) that light not resembling daylike conditions would include at least: single point sources casting sharp edged shadows (lack of scattered light from the sky), lack of or poor colour reproduction (unnatural colours), light sources off to one side only, when creating unlit hiding spots (assailant's refuge) and large deviations from similar areas (poor fit-to-area), and also lighting which concentrates light on the visibly man made objects. Yet the emphasis factors for these, in each subject's mind, are subject to personality traits, their expectations and plans, and the dominant context of use of the area.

We don't know enough to claim this proposed question as the best question, let alone final, but rather the connections between the already identified affective meters should be studied further, with larger samples.

The review also revealed contradictory hypotheses, unexpected findings, missed uncontrolled variables and a lack of holistic model for users' appraisals of lighting, or a lack of even a proposed model of appraisals for future research. Although many scientists have studied and are researching the factors behind the appraisals and how the factors and results are interconnected, there has been no multidisciplinary attempt at describing them all in one equation, or any other model.

The device and software designed, constructed and validated for functionality could be a step towards accelerating such research. It was shown to have no critical usability problems and it was shown to produce data which can be analyzed in meaningful ways.

With more test subjects and a more geographically dispersed database of appraisals and comments, further analysis algorithms for feature extraction, machine learning and pattern recognition could be used to find the determinants of good or poor locations. The appraisal dataset could even be even combined with the Mobile Laser Scanning system model of the area, if such was supplemented with accurate luminance measurements of each scene point – such an environment measurement method could be available in the future, being already researched in Aalto University Research Institute of Measuring and Modeling for the Built Environment under the Aalto University Energy Efficiency Research Programme.

The measurement of affective responses is becoming more common also outside of lighting research, and comprehensive data sets of users' appraisals should give us a better lighting and better perceived safety.

References

[1] De Boer, J. B. Visual perception in road traffic and the field of vision of the motorist. In De Boer, J. B., editor, *Public Lighting*. Eindhoven Philips Technical Library. 1967. pp. 11–96

[2] Luo Wei, Puolakka Marjukka, Viikari Meri, Kufeoglu Sinan, Ylinen Anne, Halonen Liisa. Lighting criteria for road lighting: a review. *Light & Engineering* 2012. Vol. 20 (5). pp. 64-74. Accessed 2014-08-20. Available online: http://www.sveto-tekhnika.ru/files/2012/LE42012.pdf

[3] CIE 191:2010. Recommended System for Mesopic Photometry Based on Visual Performance, Comission Internationale DE L'Eclairage CIE Central Bureau. 2010. ISBN 978-3-901906-88-6.

[4] *Tie- ja katuvalaistus.* Insinöörijärjestöjen koulutuskeskus, Suunnittelutoimikunta. Helsinki, Finland. 1984. ISBN 951-794-391-1.

[5] Kärhä, Keijo. *Katuvalaistuksen laadun ja kustannusten riippuvaisuus valaistuslaitteista*. Diplomityö. Teknillinen Korkeakoulu, Sähköteknillinen Osasto. Helsinki, Finland. 1959.

[6] Rauhala, Anne-Marjut. *Valaistus kaupunkitilan tekijänä. Jyväskylän matka pimeästä maalaiskylästä urbaanin valon kaupungiksi.* Pro gradu. Jyväskylän yliopisto. Jyväskylä, Finland. 2009. Available online `https://jyx.jyu.fi/dspace/bitstream/handle/123456789/22719/URN_NBN_fi_jyu-201001041004.pdf` Accessed 2014-07-15.

[7] Eble-Hankins, Michelle L., Waters, Clarence E. Subjective Impression of Discomfort Glare from Sources of Non-Uniform Luminance. *LEUKOS The Journal of the Illuminating Engineering Society of North America* 2009. Vol. 6 (1). pp. 51–77. DOI: DOI:10.1582/LEUKOS.2009.06.01003

[8] Lai, Daisy, Zhu, Xiaoyan, Wang, Longshi, Dermirdes Haldun, Heynderickx Ingrid. Influence of light source luminance on discomfort glare from LED road luminaires. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 220–227.

[9] Scheir, G. H., Hanselaer, P., Bracke, P., Deconinck, G., Ryckaert, W. R. Applicability of the Unified Glare Rating as Assessment of Discomfort Glare Sensation Based on Luminance Maps. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 306–313.

[10] Pong, B. J., Hsu, S. W., Chung, T. Y., Wu K. N. Simultaneous Measurements of Glare and Flicker Properties of Environmental Lightings. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 384–391.

[11] Boyce P. R. Illuminance selection based on visual performance - and other fairy stories. *Journal of Illuminating Engineering Society*. 1996. Volume 25 (2). pp. 41–49. DOI: 10.1080/00994480.1996.10748146

[12] Rautkylä, E. *Research methods and neurophysiological mechanisms behind the alerting effects of daytime light exposure*. Doctoral dissertation. Aalto University. Espoo, Finland. 2011. 103 p. ISBN 978-952-60-4249-7. Available online `https://aaltodoc.aalto.fi/handle/123456789/5017`

[13] Nikunen, Heli. *Perceptions of lighting, perceived restorativeness, preference and fear in outdoor spaces*. Ph.D. Thesis. 2013. Aalto University. Espoo, Finland. 200 p. ISBN 978-952-60-5377-6.

[14] Kuhn, L., Johansson, M., Laike, T., Govén, T. Residents' perceptions following retrofitting of residential area outdoor lighting with LEDs. *Lighting Research & Technology* 2013. Vol. 45. pp. 568–584. DOI: 10.1177/1477153512464968

[15] Unwin, J., Fotios, S. Does lighting contribute to the reassurance of pedestrians at night-time in residential roads? *Ingineria Iluminatului* 2011. Volume 13. pp. 29–44. Available online `http://journal.florinrpop.ro/28-2011-2.pdf`

[16] Davoudian, N., Raynham, P. What do pedestrians look at at night? *Lighting Research & Technology* 2012. Vol. 44. pp. 438–448.

[17] Painter, K. The influence of street lighting improvements on crime, fear and pedestrian street use, after dark. *Landscape and Urban Planning* Volume 35. Issues 2–3, August 1996, pp. 193–201.

[18] Boyce, P. R., Eklund, N. H., Hamilton, B. J., Bruno, L. D. Perceptions of safety at night in different lighting conditions. *Lighting Research & Technology* 2000. Volume 32 (2). pp. 79–91.

[19] Fotios, SA, Cheal, C, Boyce, PR. Light source spectrum, brightness perception and visual performance in pedestrian environments: a review. *Lighting Research & Technology* 2005. Volume 37. pp. 271–294.

[20] Rantakallio Antti, Nikunen Heli, Puolakka Marjukka, Halonen Liisa. *LED-ulkovalaistus kaupunkiympäristön viihtyvyyden, turvallisuuden ja energiatehokkuuden kehittämisessä - Käyttäjätutkimus*. Aalto-yliopisto Sähkötekniikan korkeakoulu, Valaistusyksikkö. 2012. Espoo, Finland. 51 p. ISBN 978-952-60-3596-3 (pdf)

[21] Knight, C. Field surveys of the effect of lamp spectrum on the perception of safety and comfort at night. *Lighting Research & Technology* 2010. Vol. 42 pp. 313–329. DOI: 10.1177/1477153510376794

[22] Raynham, P., Saksvikrønning, T. White light and facial recognition. *The Lighting Journal* 2003. Volume 68. pp. 29–33.

[23] Consolvo, S., Walker, M. Using the Experience Sampling Method to Evaluate Ubicomp Applications. *Pervasive Computing.* 2003. IEEE. Volume: 2 (2). p. 24–31. ISSN: 1536-1268. DOI: 10.1109/MPRV.2003.1203750

[24] Rohrman, B., Bishop, I. Subjective response to computer simulations of urban environments. *Journal of Environmental Psychology.* Vol. 4. pp. 319–331.

[25] Boyce, Peter Robert. *Human Factors in Lighting*, 2nd Edition. Taylor & Francis Group. 2003. ISBN 0-7484-0949-1

[26] Poulton, E. C., Quantitative subjective assessments are almost always biased, sometimes completely misleading. *British Journal of Psychology* 1977. Vol. 68. pp. 409–425.

[27] Fotios, S., Cheal, C. Obstacle detection: A pilot study investigating the effects of lamp type, illuminance and age. *Lighting Research & Technology* 2009. Volume 41: pp. 321–342.

[28] Atli, Deniz, Fotios, Steve. Rating spatial brightness: Does the number of response categories matter? *Ingineria Iluminatului* 2011. Volume 13 (1). pp. 15–28. Available online `http://journal.florinrpop.ro/27-2011-1.pdf`

[29] Fotios, Steve, Logadóttir, Ásta, Cheal, Cris, Christoffersen, Jens. Using Adjustment to Define Preferred Illuminances: Do the Results Have Any Value? *Light & Engineering* 2012. Vol. 20 (2). pp. 46–55. Accessed 2014-08-20. Available online: `http://www.sveto-tekhnika.ru/files/2012/LE22012.pdf`

[30] Fotios, S. A., Houser, K. W. Research Methods to Avoid Bias in Categorical Ratings of Brightness. *LEUKOS: The Journal of the Illuminating Engineering Society of North America* 2009. Vol. 5 (3). pp. 167–181. DOI: 10.1582/LEUKOS.2008.05.03.002

[31] Logadóttir, Á., Christoffersen, J., Fotios, S. A. Investigating the use of an adjustment task to set the preferred illuminance in a workplace environment *Lighting Research & Technology* 2011. Volume 43. pp. 403–422.

[32] Boyce, P. R, Fotios, S, Richards, M. Road lighting and energy saving. *Lighting Research & Technology* 2009. Vol. 41. pp. 245–260. DOI: 10.1177/1477153509338887

[33] Eble-Hankins, Michelle. *Subjective Impression of Discomfort Glare from Sources of Non-uniform Luminance.* 2008. PhD Dissertation. University of Nebraska, Lincoln, Nebraska. 2008.

[34] Navara, K. J., Nelson, R. J. The dark side of light at night: Physiological, epidemiological and ecological consequences. *Journal of Pineal Research* 2009. Vol. 4. pp. 215–224.

[35] Romnée, A., Lejeune, G., Bodart, M. A New Real Time Intelligent Management Model for Street Lighting. *Proceedings of CIE 2013*. Paris, France. 1239 p. ISBN 978-3-902842-44-2

[36] Tetri, E., Juntunen, E., Paakkinen, M., Tapaninen, O., Yrjänä, S., Kondratyev, V., Lehtovaara, J., Elhaddad, A. I. M., Sitomaniemi, A., Siirtola, H., Sarjanoja, E.M., Aikio, J., Halonen, L., Heikkinen, V., Nikkanen, R. AthLEDics Käyttäjän tarpeisiin vastaava energiatehokas ledivalaistus. *Aalto-yliopiston julkaisusarja Tiede + Teknologia* 1/2014. Unigrafia Oy, Helsinki. 2014. 53 p. ISBN 978-952-60-5527-5

[37] Ekrias, A., Eloholma, M., Halonen, L. Analysis of road lighting quantity and quality in varying weather conditions. *Leukos*. 2007. vol. 4. pp. 89–98.

[38] Liping, Guo. Intelligent road lighting control systems – experiences, measurements and lighting control strategies. Doctoral dissertation. *Helsinki University of Technology*. Espoo. 2008. 40 p. ISBN 978-951-22-9620-0. Available online `http://lib.tkk.fi/Diss/2008/isbn9789512296200/`

[39] Guo, Liping, Eloholma, Marjukka, Halonen, Liisa. Lighting Control Strategies for Telemanagement Road Ligting Control Systems. *LEUKOS: The Journal of the Illuminating Engineering Society of North America* 2007. Vol. 4 (3). pp. 157–171. DOI: 10.1582/LEUKOS.2007.04.03.002

[40] Smet, K. A. G. Schanda, J., Whitehead, L. CRI2012: A proposal for updating the CIE colour rendering index. *Lighting Research & Technology*. 2013. Volume 45 (6). pp. 689–709. DOI: 10.1177/1477153513481375

[41] Dangol, Rajendra. *Colour Rendering Index and colour rendering of LEDs*. Master's thesis. Aalto University School of Electrical Engineering. Espoo, Finland. 2011.

[42] Commission Internationale de l'Eclairage (CIE). *Colour rendering of White LED Light sources*. CIE Technical report 177:2007. 2007. ISBN 978-3-901906-57-2

[43] Wilkinson, R. T., Houghton, D. Field test of arousal: a portable reaction timer with data storage. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 1982. vol. 24 (4). pp. 487–493.

[44] Boomsma, Christine, Steg, Linda. Feeling Safe in the Dark: Examining the Effect of Entrapment, Lighting Levels, and Gender on Feelings of Safety and Lighting. *Environment and Behavior*. 2014. Vol. 46. (2) pp. 193–212 DOI: 10.1177/0013916512453838

[45] Rea, M. S., Bullough, J. D., Akashi, Y. A method for assessing the visibility benefits of roadway lighting. *Lighting Research & Technology* 2010. Volume 42: pp. 215–239. DOI: 10.1177/1477153509360855

[46] Bishop, I. D., Rohrmann, B., Subjective responses to simulated and real environments: a comparison. *Landscape and Urban Planning.* 2003. Volume 65. p. 261–277. DOI: 10.1016/S0169-2046(03)00070-7

[47] Houtkamp, Joske M, Toet, Alexander. Who's Afraid of Virtual Darkness - Affective Appraisal of Night-time Virtual Environments. *Proceedings of digital landscape architecture 2012.* Herbert Wilchmann Verlag. Berlin, Germany. 2012. pp. 508–515. ISBN 9783879075195.

[48] De Kort, Y. A. W., Meijnders, A. L., Sponselee, A. A. G., Ijsselsteijn, W. A. What's wrong with virtual trees? Restoring from stress in a mediated environment. *Journal of Environmental Psychology.* 2006. Vol. 26. pp. 309–320.

[49] Newsham, G. R., Richardson, C., Blanchet, C., Veitch, JA. Lighting quality research using rendered images of offices. *Lighting Research and Technology* 2005. Volume 37 (2). pp. 93–115.

[50] Eastman, C., Young, M., Fogg, L., Liu, L., Meaden, P. Bright Light Treatment of Winter Depression A Placebo-Controlled Trial. *Archives of General Psychiatry.* 1998. Volume 55(10). pp. 883–889. DOI:10.1001/archpsyc.55.10.883

[51] SFS-EN 13201-1:2004. Road lighting. Part 1: Selection of lighting classes.

[52] Tiehallinto. *Tievalaistuksen suunnittelu.* Edita Prima Oy, Helsinki. 2006. 115 p. Accessed 2014-08-14. Available online: `http://alk.tiehallinto.fi/thohje/pdf/2100034-v-06tievalaist_suunn.pdf`

[53] Tang, Tingan. *Combining User and Context: Living Labs Innovation in Digital Services* Ph.D. Thesis. Aalto University. Espoo, Finland. 2014. ISBN 978-952-60-5741-5.

[54] Raymond, Eric S. *The cathedral and the bazaar : musings on Linux and open source by an accidental revolutionary.* O'Reilly & Associates, Inc. Sebastopol, CA. 2001. 241 p. ISBN 0-596-00108-8. Accessed 2014-08-18. Available online: `http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/`

[55] Guzman, J. G., del Carpio, A. F., Colomo-Palacios, R., de Diego, M. V. Living Labs for User-Driven Innovation A Process Reference Model. *Research Technology Management* 2013. Vol. 56 (3). pp. 29–39. DOI: 10.5437/08956308X5603087

[56] Haans, A., de Kort, Y. A. W. Light distribution in dynamic street lighting: Two experimental studies on its effects on perceived safety, concealment, and escape. *Journal of Environmental Psychology* 2012. Vol. 32. pp. 342–352.

[57] Howe, J. The Rise of Crowdsourcing. *Wired.* 2006. Vol. 14 (6). Accessed 18.6.2014. Available online `http://www.wired.com/wired/archive/14.06/crowds_pr.html`

[58] Surowiecki, J. *The Wisdom of Crowds*. Random House LLC. 2005. 296 p. ISBN: 0385721706

[59] Brabham, Daren C. Crowdsourcing as a Model for Problem Solving An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*. Vol. 14(1). pp. 75–90. DOI: 10.1177/1354856507084420

[60] Balfour, Jennifer L, Kaplan, George A. Neighborhood Environment and Loss of Physical Function in Older Adults: Evidence from the Alameda County Study. *American Journal of Epidemiology* Vol. 155 (6). pp. 507–515

[61] Russel, J. A., Lanius, U. F. Adaptation Level and the Affective Appraisals of Environments. *Journal of Environmental Psychology* 1984. Vol. 4. pp. 119–135.

[62] Hektner, Joel M., Schmidt, Jennifer A., Csikszentmihalyi, Mihaly. *Experience Sampling Method*. Thousand Oaks, CA. SAGE Publications, Inc. 2007. SAGE Research Methods. DOI: 10.4135/9781412984201

[63] Mäkelä Olli, Kärki Jutta-Leea. *Tievalaistuksen suunnittelu*. Tiehallinnon selvityksiä 18/2004. Edita Prima Oy, Helsinki, Finland. 2004. 53 p. Accessed 2014-08-15. Available online: `http://alk.tiehallinto.fi/julkaisut/pdf/3200868-vtievalvaik.pdf`

[64] Saastamoinen Kimmo, Kärki Jutta-Leea, Mäkelä Olli. *Ajonopeudet taajamissa* Tiehallinnon selvityksiä 2/2003. Helsinki, Finland. 2003. Accessed 2014-08-15. Available online: `http://alk.tiehallinto.fi/julkaisut/pdf/3200791ajonopeudettaajamissa.pdf`

[65] Nielsen, Jakob. *Usability engineering*. Academic Press. Boston, USA. 1993. 362 p. ISBN 0-12-518405-0, ISBN 0-12-518406-9

[66] International Standards Organisation. *ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*. 1998.

[67] Riihiaho, S. The Pluralistic Usability Walk-Through Method. *Ergonomics in Design* 2002. Vol. 10 (3). pp. 23–30. Available online: `http://www.soberit.hut.fi/~sri/pluralistic.pdf` Accessed 2014-07-15.

[68] Teague, Ross, De Jesus, Katherine, Nunes-Ueno, Marcos. Concurrent Vs. Post-Task Usability Test Ratings. in *CHI 2001 Extended Abstracts on human factors in computing systems*. 2001. pp. 289–290. DOI: 10.1145/634067.634238

[69] Wanvik Per Ole. Effects of road lighting: An analysis based on Dutch accident statistics 1987–2006. *Accident Analysis and Prevention* 2009. Volume 41. pp. 123–128.

[70] Garg, K., Nayar, S., Vision and rain. *International Journal of Computer Vision* 2007. Vol. 75 (1). pp. 3–27. DOI: 10.1007/s11263-006-0028-6

[71] Bernardin, Frédéric, Bremond, Roland, Ledoux Vincent, Pinto Maria, Lemonnier Sophie, Cavallo, Viola, Colomb, Michèle. Measuring the effect of the rainfall on the windshield in terms of visual performance. *Accident Analysis & Prevention* Volume 63. pp. 83–88. ISSN 0001-4575. DOI: 10.1016/j.aap.2013.10.008

[72] OECD/ITF (editors) *Road Safety Annual Report 2014*. Paris: International Traffic Safety Data and Analysis Group, International Transport Forum. 2014. 526 p. Available online: `http://www.internationaltransportforum.org/pub/pdf/14IrtadReport.pdf` Accessed 2014-05-30.

[73] Fastenmeier, Wolfgang. Driving task analysis as a tool in traffic safety research and practice *Safety Science*. 2007. Vol. 45(9). pp. 952–979

[74] Kramer, A., Wiegmann, D., Kirlik, A., *Attention : From Theory to Practice*. Oxford, GBR: Oxford University Press, USA. 2006. ISBN 9780195305722

[75] Deery, Hamish A. Hazard and Risk Perception among Young Novice Drivers. *Journal of Safety Research*. 1999. Volume 30 (4). p. 225–236. DOI: 10.1016/S0022-4375(99)00018-3

[76] Fastenmeier, W. Die Verkehrssituation als Analyseeinheit im Verkehrssystem (The road traffic situation as analysis unit in the road traffic system). In: Fastenmeier, W. (Ed.), *Autofahrer und Verkehrssituation - Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme*. Verlag TÜV Rheinland, Köln. 1995. pp. 27–78.

[77] Blöbaum, A., Hunecke, M. Perceived danger in urban public space. The impacts of physical features and personal factors. *Environment and Behavior* 2005. Vol. 37 (4). pp. 465–486. DOI: 10.1177/0013916504269643

[78] Viliunas, V., Vaitkevicius, H., Stanikunas, R., Vitta, P., Bliumas, R., Auskalnyte, A., Tuzikas, A., Petrulis, A., Dabasinskas, L., Zukauskas, A. Subjective evaluation of luminance distribution for intelligent outdoor lighting. *Lighting Research and Technology* 2014. Vol. 46. pp. 421–433. DOI: 10.1177/1477153513491760

[79] Atkins, Stephen, Husain, Sohail, Storey, Angele. *The Influence of Street Lighting on Crime & Fear of Crime*. Paper 28, Crime Prevention Unit. London. 1991. 59 p. ISBN: 0-86252-668-X

[80] Fotios, S., Yang, B., Uttley, J. Observing other pedestrians: Investigating the typical distance and duration of fixation. *Lighting Research and Technology* published online 3 April 2014. 17 p. DOI: 10.1177/1477153514529299

[81] Alferdinck, J. W. A. M., Hogervorst, M. A., van Eijk, A. M. J., Kusmierczyk, J. T. *Mesopic vision and public lighting - A literature review and a face recognition experiment* (TNO-report TNO-DV 2010 C435). TNO Defence, Security and Safety. Soesterberg, The Netherlands. 2010. 89 p.

[82] Akashi, Y., Rea, M. S., Bullough, J. D. Driver decision making in response to peripheral moving targets under mesopic light levels. *Lighting Research & Technology* 2007. Volume 39, pp. 53–66.

[83] Eloholma, Marjukka. *Development of visual performance based mesopic photometry.* Ph.D. Thesis. Helsinki University of Technology. Espoo, Finland. 2005. ISBN 951-22-7868-5. Accessed 2014-03-17. Available online `http://lib.tkk.fi/Diss/2005/isbn9512278685/isbn9512278685.pdf`

[84] Iwata, Michico, Ayama, Miyoshi et al. Visibility evaluation for face of person standing under street lighting environment. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency.* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 538–545.

[85] Fotios, S., Uttley, J., Yang, B. Lighting for pedestrians: what are the critical visual tasks? *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency.* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 164–173. ISBN 978-3-902842-49-7.

[86] Rea, M. S., Bullough, J. D., Akashi, Y. Several views of metal halide and high-pressure sodium lighting for outdoor applications. *Lighting Research & Technology* 2009. Volume 41: pp. 297–320.

[87] Rombauts, Patrick, Vandewyngaerde, Hans, Maggetto, Gaston. Minimum semi-cylindrical illuminance and modelling in residential area lighting. *Lighting Research & Technology* 1989. Vol. 21 (2). pp. 49–55.

[88] Johansson, M., Rosén, M., Küller, R. Individual factors influencing the assessment of the outdoor lighting of an urban footpath. *Lighting Research and Technology* 2011. Volume 43(1). pp. 31–43.

[89] Fotios, S., Unwin, J., Road lighting and pedestrian reassurance after dark: A review. *Lighting Research & Technology* published online 21 February 2014. 21 p.

[90] Jaatinen, Saara. *The use of LEDs in pedestrian ways, parking areas and parks - A user study.* Master's thesis. Aalto University School of Electrical Engineering. Espoo, Finland. 2010. Accessed 2014-03-10. Available online: `http://lib.tkk.fi/Dipl/2010/urn100269.pdf`

[91] Cengiz, C., Kotkanen, H., Puolakka, M., Lappi, O., Lehtonen, E., Halonen L., Summala, H. Combined eye-tracking and luminance measurements while driving on a rural road: Towards determining mesopic adaptation luminance.

Lighting Research & Technology published online 26 September 2013. 19 p. DOI: 10.1177/1477153513503361

[92] Uchida, T., Ohno, Y. Defining the visual adaptation field for mesopic photometry: Does surrounding luminance affect peripheral adaptation? *Lighting Research and Technology.* 2014. Vol. 46. pp. 520–533. DOI: 10.1177/1477153513498084

[93] Rantakallio, Antti. *Suomen tieverkko, nykyaikaiset valaistusvaihtoehdot ja käyttäjätutkimus ledivalaistukselle.* Diplomityö. Aalto-yliopisto Sähkötekniikan Korkeakoulu, Espoo, Finland. 2011. 100 p.

[94] Caberletti, L., Elfmann, K., Kümmel, M., Schierz, C. Influence of Ambient Lighting in Vehicle Interior on the Driver!s Perception. In: *Proceedings of Experiencing Light 2009 International Conference on the Effects of Light on Wellbeing* (Eds. Y. A. W. de Kort, W. A. IJsselsteijn, I. M. L. C. Vogels, M. P. J. Aarts, A. D. Tenner, and K. C. H. J. Smolders). 2009. pp. 5–13.

[95] Akashi, Y., Rea, M. S., Bullough, J. D. Predicting discomfort glare from outdoor lighting installations. *Lighting Research and Technology* 2008. Volume 40. pp. 225–238. DOI: 10.1177/1477153508094048.

[96] Ekrias, Aleksanteri, Eloholma, Marjukka, Halonen, Liisa. The effects of colour contrast and pavement aggregate type on road lighting performance. *Light & Engineering* 2009. vol. 17, pp. 76–91.

[97] Viikari M, Puolakka M, Halonen L, Rantakallio A. Road lighting in change: User advice for designers. *Lighting Research & Technology* 2012. Vol. 44. pp. 171–185. DOI: 10.1177/1477153511424473

[98] Johansson, M., Pedersen, E., Maleetipwan-Mattsson, P., Kuhn, L., Laike, T. Perceived outdoor lighting quality (POLQ): A lighting assessment tool. *Journal of Environmental Psychology* 2014. Vol. 39. pp. 14–21. DOI: 10.1016/j.jenvp.2013.12.002

[99] Romnée, A., Bodart, M. Street lighting appreciated by pedestrians: a field study. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 862–872. ISBN 978-3-902842-49-7.

[100] Kostic, A., Djokic, L. Subjective impressions under LED and metal halide lighting. *Lighting Research & Technology* 2014. Vol. 46. pp. 293–307. DOI: 10.1177/1477153513481037

[101] Juntunen, E., Tetri, E., Tapaninen, O., Yrjänä, S., Kondratyev, V., Sitomaniemi, A., Siirtola H., Sarjanoja, E. M., Aikio, J., Heikkinen, V. A smart LED luminaire for energy savings in pedestrian road lighting. *Lighting Research & Technology.* First published online 7 November 2013. DOI: 10.1177/1477153513510015

[102] Poole, Alex, Ball, Linden J. Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future. in C. Ghaoui (Ed.): *Encyclopedia of Human-Computer Interaction.* Idea Group, Inc. Pennsylvania. 2005.

[103] Patla, Aftab E., Vickers, Joan N. Where and when do we look as we approach and step over an obstacle in the travel path? *NeuroReport* 1997. Vol. 8 (17). pp. 3661–3665.

[104] Fujiyama, Taku, Childs, Graig, Boampong, Derrick, Tyler, Nick. *How do elderly pedestrians perceive hazards in the street?* Transport Canada. 2007. Accessed 2014-08-19. Available online: `http://eprints.ucl.ac.uk/5030/`

[105] Mack, Arien. Inattentional Blindness: Looking Without Seeing *Current Directions in Psychological Science* 2003. Vol. 12 (5). pp. 180–184.

[106] O'Regan, J. Kevin, Deubel, Heiner, Clark, James J., Rensink, Ronald A. Picture Changes During Blinks: Looking Without Seeing and Seeing Without Looking. *Visual Cognition* 2000. Vol. 7. pp. 191–211

[107] Pappas, Jennifer M., Hicks, Jacob M. An eye-tracking approach to inattentional blindness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2005 Vol. 49 (17). pp. 1658–1662.

[108] Luo, Wei. *Outdoor lighting - Mesopic photometry, adaptation conditions and user preferences in pedestrian way lighting.* Doctoral dissertation. *Aalto University.* Espoo. 2014. 66 p. ISBN 978-952-60-5723-1 (pdf). Available online `http://urn.fi/URN:ISBN:978-952-60-5723-1` Accessed 2014-09-25

[109] Fotios, S., Dong, M., Yang, B., Lin, Y. Interpersonal judgements, lamp spectrum and task difficulty. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency.* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 357–366. ISBN 978-3-902842-49-7

[110] Figueiro, M., Plitnick, B., Rea, M., Gras, L. Rea, M. Lighting and perceptual cues: Effects on gait measures of older adults at high and low risk for falls. *BMC Geriatrics.* 2011. Vol. 11. p. 49.

[111] Hakkarainen, Henri. *Ledien käyttömahdollisuudet tievalaistuksessa.* Diplomityö. Tampereen Teknillinen Yliopisto. 2010. 71 p.

[112] *Myllyojan katuvalaistuksen testikentän raportti.* Oulun Energia Urakointi Oy. 2012. Available online `http://oulu.ouka.fi/tekninen/suunnitelmat/Nayta_Liite.asp?ID=3049&Liite=Myllyojan%20katuvalaistuksen%20testikent%E4n%20raportti.pdf`

[113] Royer, M. P., Poplawski, M. E., Tuenge, J. R. *Demonstration Assessment of LED Roadway Lighting.* Pacific Northwest National Laboratory report. Washington, USA. 2012. 67 p. Available online: `http://apps1.eere.energy.gov/`

buildings/publications/pdfs/ssl/2012_gateway_cully.pdf Accessed 2014-08-13.

[114] Vilmi, Toivo. Käyttäjäkokemukset valaisinvalmistajan näkökulmasta – saadut opit. *Athledics seminar* 2012. Available online http://www.lightinglab.fi/athledics/News/workshop/10-vilmi-kayttajakokemukset%20valaisinvalmistajan.pdf Accessed 2014-08-13.

[115] Yang, B., Fotios, S. Lighting and recognition of emotion conveyed by facial expressions. *Lighting Research & Technology.* 2014. First published online 11 September 2014. DOI: 10.1177/1477153514547753

[116] Veitch, Jennifer A. Psychological Processes Influencing Lighting Quality. *Journal of the Illuminating Engineering Society* 2001. Vol. 30 (1). pp. 124–140.

[117] Hanyu, K. Visual properties and affective appraisals in residential areas after dark. *Journal of Environmental Psychology* 1997. Vol. 17. pp. 301–315.

[118] Faulkner, Xristine. *Usability Engineering.* Palgrave, New York, NY, USA. 2000. 244 p. ISBN 0-333-77321-7

[119] Govén, T., Gentile, N., Laike, T., Sjöberg, K. Energy efficient and study promoting lighting at high school: preliminary results. *Proceedings of CIE 2014 Lighting Quality and Energy Efficiency* x039:2014. Kuala Lumpur, Malaysia. 2014. pp. 772–779.

[120] Zak, Pavel P., Ostrovsky, Mikhail A. Potential danger of light emitting diode illumination to the eye, in children and teenagers. *Light & Engineering* 2012. Vol. 20 (3). pp. 5–8. Available online: http://www.sveto-tekhnika.ru/files/2012/LE32012.pdf Accessed 2014-08-20.

[121] Falchi, F., Cinzano, P., Elvidge, C. D., Keith, D. M., Haim, A. Limiting the impact of light pollution on human health, environment and stellar visibility. *Journal of Environmental Management* 2011. Vol 92. pp. 2714–2722.

[122] Hur, M., Nasar, J. L., Chun, B. Neighborhood satisfaction, physical and perceived naturalness and openness. *Journal of Environmental Psychology* 2010. Vol. 30. pp. 52–59. DOI: 10.1016/j.jenvp.2009.05.005

[123] Craglia, Max. Multidisciplinary Research on Geographical Information in Europe and Beyond. In: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke (Editors) *Proceedings of the AGILE'2012 International Conference on Geographic Information Science.* Avignon, France. 2012. ISBN: 978-90-816960-0-5

[124] Soneira, Raymond M., *Why Existing Brightness Controls and Light Sensors are Effectively Useless* 2010. Available online http://www.displaymate.com/AutoBrightness_Controls_2.htm Accessed 2014-08-12.

[125] Kimpe, Tom, Tuytschaever, Tom. Increasing the Number of Gray Shades in Medical Display Systems – How Much is Enough? *Journal of Digital Imaging* 2007. Volume 20 (4). pp. 422–432. DOI: 10.1007/s10278-006-1052-3

[126] Reijula, Jori. *Using well-being technology in monitoring elderly people - a new service concept* Doctoral Dissertation. Aalto University School of Science and Technology. Espoo, Finland. 2010. ISBN (pdf) 978-952-60-3309-9. Available online: `http://lib.tkk.fi/Diss/2010/isbn9789526033099/` Accessed 2014-03-04.

[127] Android developers Dashboard. `http://developer.android.com/about/dashboards/index.html` Accessed October 7 2014.

[128] Nielsen, Jakob, Landauer, Thomas K. A mathematical model of the finding of usability problems. In: *Proceedings of ACM INTERCHI'93 Conference*. Amsterdam, The Netherlands 24-29 April 1993. pp. 206-213. DOI: 10.1145/169059.169166

A Input device construction details

The schematic for the input device's electronics does not present anything beyond simple connections, but serves to point out that the planning and contruction of the device's electronics was straightforward. Relative more effort was spent on the physical container, finding it and altering it, and on finding the suitable tactive buttons for the user interface. Implementing the Android software required considerably more work than the device.

The open source software Fritzing was used for drawing the electronic schematic and the layout diagram below.

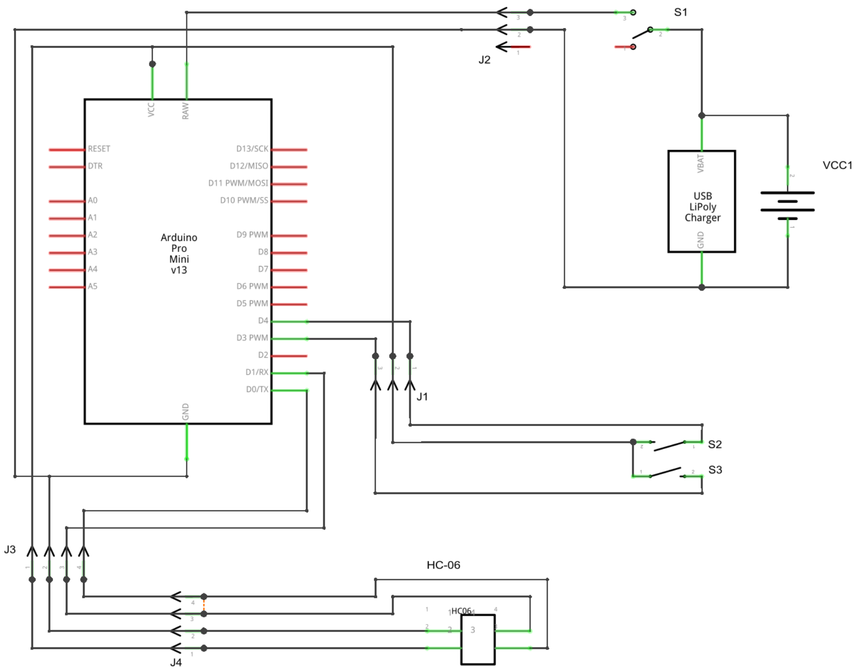

Figure 10: Electronic schematic of the feedback device

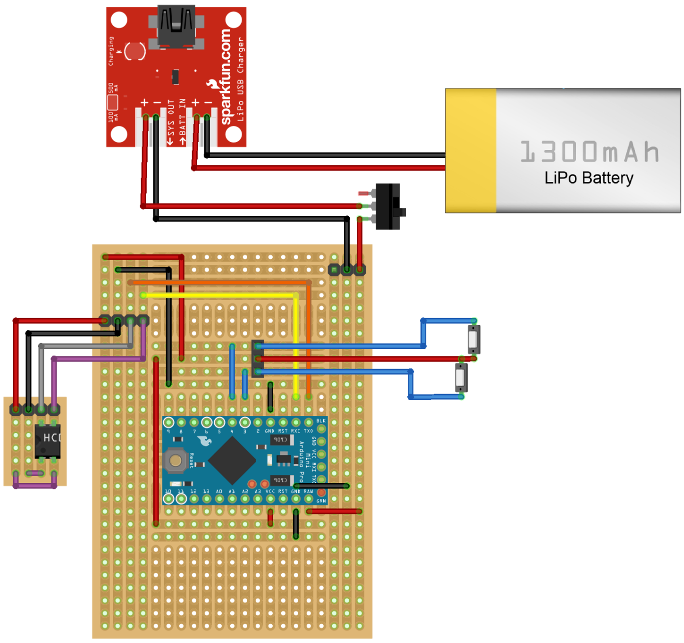

Figure 11: Layout diagram of the feedback device

Table 9: Input device build costs.

10,00	Arduino Pro Mini 3.3v
13,00	Seeedstudio HC-06 Brick
14,00	Sparkfun USB LiPoly Charger - Single Cell
9,90	Polymer Lithium Ion Battery 1400 mAh
2,00	Perforated stripboard
0,50	2x button
	wires
2,50	enclosure
51,90	total (EUR)

B Input device source code

Basically the code for the device's microcontroller constantly checks if there's input to read from the Bluetooth link, and if not, checks if the state of any of the buttons has changed since the last loop iteration. Button state changes are filtered by first confirming they stay in their new state for at least 50 milliseconds, before sending an event text to the Bluetooth connection. This is necessary, because the state of the button inputs fluctuates between pressed and not pressed several times while the physical button is moving down, or up.

The source code for the Android software was too long to include as an appendix, and uninteresting from the academic point of view. Listing the 309 kilobytes of source code would have required approximately 170 pages.

```
int led = 13;
int ledstate = LOW;
int buttonPin = 3;
int button2Pin = 4;
int buttonState = HIGH;
int button2State = HIGH;
int lastButtonState = HIGH;
int lastButton2State = HIGH;
int reading = LOW;
int reading2 = LOW;
// when switching between loops, we want to reply the whole string:
int firstReadByte = 0;
int sentLastButton = 0;  // after debounce, only send once
int sentLastButton2 = 0;  // after debounce, only send once
long lastDebounceTime = 0;
long previousChangeTime = 0;
long lastDebounceTime2 = 0;
long previousChangeTime2 = 0;
long debounceDelay = 50;
long msgRcvTime = 0;
// difference between ms time from mobile and
// this local ms, at last message from the other end:
long timediff = 0;

void setup() {
  pinMode(led, OUTPUT); // internal led of Pro Mini
  pinMode(buttonPin, INPUT_PULLUP);
  pinMode(button2Pin, INPUT_PULLUP);
  Serial.begin(9600); //note this may need to be changed to match your module
  // Most likely the BT hasn't paired yet, so these are never received
  Serial.println("Calling Ligappr,20141016...");
  Serial.println("MYTIMEEND");
}
```

```
void loop() {
  while(Serial.available()==0) {
    // when there’s no incoming data from BT, read buttons:
    reading = digitalRead(buttonPin);
    reading2 = digitalRead(button2Pin);
    // buttons’ states fluctuate when they’re getting pressed or
    // released, so there has to be filtering; "debouncing"
    if (reading != buttonState) {
      // if it diverges from current "confirmed" state, examine interval:
      // timestamp of latest change of input state
      lastDebounceTime = millis();
      if ( (lastDebounceTime - previousChangeTime ) > debounceDelay) {
        // enough time had passed from last confirmed change
        buttonState = reading;
        previousChangeTime = lastDebounceTime;
      }
    }
    // internal pull-up resistor, so pushed -> LOW
    if (buttonState == LOW) {
      // turn tiny internal LED on for diagnostics:
      digitalWrite(led, HIGH);
      if (sentLastButton == 0) {
        // write time and button to BT:
        Serial.write("At(");
        Serial.print(previousChangeTime);
        Serial.print(")Button1");
        Serial.println("END");
        // don’t send this key press again:
        sentLastButton = 1;
      }
    } else {
      digitalWrite(led, LOW);
      sentLastButton = 0; // can send next button
    }
    if (reading2 != button2State) {
      // if it diverges from the "confirmed" state, examine interval:
      // timestamp of latest change of input state
      lastDebounceTime2 = millis();
      if ( (lastDebounceTime2 - previousChangeTime2 ) > debounceDelay) {
        // enough time had passed from confirmed change
        button2State = reading2;
        previousChangeTime2 = lastDebounceTime2;
      }
    }
```

```
    if (button2State == LOW) {
      if (sentLastButton2 == 0) {
        Serial.write("At(");
        Serial.print(previousChangeTime2);
        Serial.print(")Button2");
        Serial.println("END");
        sentLastButton2 = 1; // don't send this again
      }
    } else {
      sentLastButton2 = 0; // can send next button
    }
  }
  // when the while loop above exits, there's something to read on serial:
  msgRcvTime = millis();
  firstReadByte = 0; // serial buffer was empty
  lastButtonState = reading;
  // reply the message for ACK
  Serial.write("At(");
  Serial.print(msgRcvTime);
  Serial.write(")I heard you say:");
  // read everything incoming
  while(Serial.available()>0) {
    Serial.write(Serial.read());// note it is Serial.WRITE
    // wait 1 millisecond once, so that we don't reply just
    // the first byte; by then there should be more bytes
    // in the buffer. At baud rate 9600, one byte takes less
    // than 1 ms.
    if (firstReadByte == 0) {
      delay(1);
      firstReadByte = 1;
    }
  }
  // End-of-message token for the other end to parse
  Serial.println("END");
}
```